

**UNIVERSITÀ DEGLI STUDI DI PADOVA**

**FACOLTÀ DI SCIENZE STATISTICHE**

Corso di Laurea in Scienze Statistiche, Demografiche e Sociali



**TESI DI LAUREA**

**LA CONCENTRAZIONE DI INQUINANTI  
IN POZZI DELL'ALTA PADOVANA:  
UN'ANALISI MEDIANTE IL *KRIGING***

Relatore: Dott. Carlo Gaetan

Correlatore: Dott. Paolo Fabbri

Laureanda: Monica Lorenzetto

Anno Accademico 2002/2003



# Indice

<b>1</b>	<b>Le acque sotterranee</b>	<b>1</b>
1.1	Le acque sotterranee nella pianura veneta . . . . .	3
1.2	Inquinanti delle acque sotterranee . . . . .	6
1.3	Sorgenti di inquinamento . . . . .	6
1.4	Movimento di un inquinante nel sottosuolo . . . . .	7
<b>2</b>	<b>Modellazione geostatistica</b>	<b>9</b>
2.1	Introduzione . . . . .	9
2.2	Le variabili regionalizzate . . . . .	10
2.3	Proprietà del variogramma . . . . .	13
2.3.1	Isotropia e anisotropia . . . . .	14
2.3.2	I modelli teorici . . . . .	16
2.4	Il variogramma empirico . . . . .	19
2.4.1	Nuvola del variogramma . . . . .	20
2.5	Aspetti inferenziali . . . . .	21
2.6	Il <i>kriging</i> . . . . .	22
2.7	<i>Kriging</i> ordinario . . . . .	23
2.7.1	<i>Kriging</i> puntuale . . . . .	23
2.8	Metodi non stazionari . . . . .	25
2.8.1	<i>Kriging</i> universale . . . . .	25
2.9	Convalida incrociata . . . . .	27
<b>3</b>	<b>Analisi geostatistica</b>	<b>29</b>
<b>4</b>	<b>Applicazioni</b>	<b>35</b>
4.1	Introduzione . . . . .	35
4.2	I Cloruri . . . . .	38
4.2.1	Descrizione e rappresentazione spaziale dei dati . . . . .	38

4.2.2	Stima del variogramma . . . . .	40
4.2.3	Costruzione del modello - Stima dei parametri . . . . .	43
4.2.4	Previsione e rappresentazione grafica della concentrazione di Cloruri . . . . .	45
4.2.5	Convalida del modello . . . . .	48
4.3	I Nitrati . . . . .	48
4.3.1	Descrizione dei dati . . . . .	48
4.3.2	Stima del variogramma . . . . .	50
4.3.3	Costruzione del modello . . . . .	52
4.3.4	Previsione e rappresentazione grafica della concentrazione di Nitrati . . . . .	53
4.3.5	Convalida del modello . . . . .	56
4.4	I Composti Organo-alogenati . . . . .	56
4.4.1	Descrizione dei dati . . . . .	56
4.4.2	Stima del variogramma . . . . .	59
4.4.3	Costruzione del modello . . . . .	60
4.4.4	Previsione e rappresentazione grafica della concentrazione di Composti organo-alogenati . . . . .	61
4.4.5	Convalida del modello . . . . .	63
<b>A</b>	<b>Rete di monitoraggio</b>	<b>67</b>
<b>B</b>	<b>I dati analizzati</b>	<b>69</b>
B.1	I 68 pozzi con profondità minore di 30 metri . . . . .	70
B.2	I 49 pozzi con profondità compresa tra 30 e 60 metri . . . . .	72
<b>C</b>	<b>Zona di interesse e stratigrafie</b>	<b>75</b>
<b>D</b>	<b>Il software R</b>	<b>81</b>



# Introduzione

Questa tesi è nata con lo scopo di fornire una metodologia per la modellazione e la previsione spaziale della concentrazione di inquinanti in acque sotterranee, a partire da misure effettuate in pozzi campionati.

Spesso chi si occupa di indagare su eventuali situazioni di degrado idrico possiede un'accurata conoscenza del fenomeno, dal punto di vista deterministico, dato dalle leggi chimiche, fisiche e geologiche che lo contraddistinguono, ma non degli strumenti adatti a trattare una grande mole di dati. Per questo motivo si è deciso di studiare una base dati fornita dall'ARPAV, integrando le conoscenze statistiche con quelle della geologia. Dunque la tesi non vuole tanto essere uno studio della concentrazione di alcuni inquinanti, ma piuttosto un'applicazione di una metodologia riproducibile per altri insiemi di dati.

Attraverso l'utilizzo di un potente *software*, quale è il linguaggio R (Versione 1.7.1, Appendice D), è stato possibile condurre l'intera analisi, dall'esplorazione preliminare alla previsione spaziale di specifici inquinanti a determinate profondità. Con altri *software* è, invece, necessario spezzare le fasi di studio, importando i dati da un ambiente all'altro, e quindi vengono enormemente dilatati i tempi di raggiungimento dei risultati.

Il primo passo dell'analisi è stato l'esplorazione preliminare delle 5862 osservazioni effettuate in pozzi di 15 comuni dell'Alta Padovana (Cittadella, Tombolo, San Pietro in Gù, ecc). Ciascuna delle osservazioni consisteva nelle coordinate UTM del pozzo, il comune, la profondità, la data del prelievo e una serie di variabili chimiche e fisiche caratterizzanti i campioni di acqua. Successivamente i risultati delle statistiche descrittive sono stati integrati con un'analisi stratigrafica della zona di interesse, che ha condotto ad una suddivisione della base dati in insiemi omogenei secondo la profondità (minore di 30 metri, tra 30 e 60 metri, maggiore di 60 metri). Un'ulteriore semplificazione si è avuta dividendo le osservazioni in base all'anno di rilevazione (dal 1990 al 1996). In questo modo si sono ottenuti insiemi contenenti un numero limitato di os-

servazioni per le variabili su cui effettuare la previsione spaziale, dopo averle depurate dei dati mancanti e delle ridondanze.

I dati che riguardano le scienze della terra sono variabili distribuite nello spazio e nel tempo. La conoscenza della concentrazione di un inquinante, della temperatura o della piovosità non rivestono particolare interesse se non sono noti i siti e i periodi in cui sono state effettuate le misurazioni. La Geostatistica fornisce gli strumenti statistici per affrontare le situazioni in cui è necessario tenere conto della dipendenza spaziale e temporale delle osservazioni.

Lo sviluppo della geostatistica è iniziato negli anni '50 del secolo scorso quando, in Sud Africa, D. Krige cominciò ad applicare tecniche statistiche per la previsione delle riserve minerarie. A partire dagli anni '60 il matematico francese G. Matheron formulò le basi teoriche di tali metodi e un modello che tenesse conto delle fluttuazioni delle variabili osservate ed, in particolar modo, della correlazione spaziale. Matheron (1971) ha definito le variabili spaziali come variabili regionalizzate per descrivere la loro natura, espressa da due aspetti complementari: uno casuale, che tiene conto di irregolarità locali presentate dal fenomeno (*piccola scala*), e un aspetto strutturale, che riflette le caratteristiche globali dello stesso (*larga scala*). Le variabili regionalizzate sono la realizzazione di processi stocastici, e uno degli obiettivi della geostatistica è quello di identificare le due componenti che le caratterizzano, a partire dalle informazioni fornite dal campione osservato. Nella maggior parte dei casi è impossibile controllare l'assunto che una variabile sia la realizzazione di una funzione casuale, avendo a che fare con una singola realizzazione. E' necessario allora imporre delle ipotesi di stazionarietà, tramite le quali è possibile ritenere il fenomeno omogeneo, e si può definire la funzione *variogramma* che in geostatistica descrive la correlazione spaziale.

Dagli anni '80 la geostatistica è un mezzo per descrivere la variabilità spaziale e l'applicazione più diffusa, il *kriging*, consiste nella previsione dei valori di una grandezza in punti in cui non è possibile effettuare misurazioni. Tale obiettivo si può raggiungere utilizzando le osservazioni disponibili nei punti campionati, supponendo l'esistenza di una relazione tra valori osservati e incogniti. Proprio per questo la Geostatistica è stata inizialmente utilizzata dall'industria mineraria per ridurre gli elevati costi delle perforazioni. In seguito si sono rivelati i molteplici campi di applicazione di tale disciplina: scienze del suolo, idrogeologia, meteorologia, scienze ambientali, agricoltura, ingegneria civile, ecc.

Nella tesi si presenta inizialmente una panoramica sul fenomeno oggetto di studio. Nel Capitolo 1, infatti, vengono descritti i tipi di pozzi che si possono trovare nella Pianura Veneta e in particolare nella zona di interesse (viene inquadrata nel Capitolo 3), in base alle caratteristiche delle acque sotterranee (falde, acquiferi). Inoltre sono presentati i tipi di inquinamento e le principali sostanze chimiche potenzialmente inquinanti per le acque sotterranee.

Le applicazioni del Capitolo 4 si riferiscono a misure di concentrazione dei principali inquinanti delle acque sotterranee quali Cloruri, Nitrati (come  $NO_3$ ) e Composti Organo-alogenati effettuate nel 1996 (dal 04/01/1996 al 22/08/1996) in pozzi di profondità minore di 30 metri e compresa tra 30 e 60 metri. I metodi utilizzati nelle applicazioni sono analisi statistiche descrittive, inferenziali e geostatistiche specifiche per ricerche di questo tipo, in particolare il *kriging*. Nel secondo Capitolo vengono trattati i fondamenti teorici di questa metodologia.

Le procedure utilizzate per l'analisi geostatistica coinvolgono sia tecniche puramente statistiche, sia procedimenti cartografici per mezzo dei quali si ha un riscontro visivo dei risultati matematici ottenuti e si può comprendere in modo migliore il comportamento del fenomeno studiato. Spesso si hanno a disposizione quantità limitate di osservazioni, riguardo ad una variabile di interesse, con le quali costruire le mappe di previsione, e perciò queste avranno un basso grado di risoluzione e un'elevata varianza di previsione. In situazioni di questo tipo si ricorre a informazioni supplementari quali possono essere ad esempio la stratigrafia (Appendice C) o la conoscenza delle attività umane svolte nella zona di interesse, nel caso si eseguano studi di idrogeologia.

I risultati sono mappe di concentrazione di nitrati, composti organo-alogenati e cloruri, estese a tutta l'area in esame, ottenute tramite il *kriging*. Si tratta di una media mobile pesata, che permette di effettuare previsioni non distorte, su osservazioni localizzate nei pressi dei punti non campionati, attraverso un predittore lineare, ricavato per interpolazione dalle rilevazioni eseguite sui pozzi campionati nel 1996. La conclusione dell'analisi geostatistica avviene appunto con la restituzione cartografica dei valori dell'attributo studiato nella regione indagata, ottenuti appunto tramite il *kriging*.

Nel Capitolo 3 sono presentati i passi che hanno portato alla realizzazione delle mappe, elaborate mediante il *software* R.

A conclusione della tesi si è osservato che la modellazione e previsione spaziale sono utili per indagare eventuali situazioni di degrado nelle acque. La rappresentazione delle mappe di concentrazione ha permesso di notare come

i cloruri presenti abbiano origine naturale, viste le basse concentrazioni. Lo ione nitrato, ritenuto un parametro significativo per quanto riguarda l'inquinamento diffuso, è risultato avere valori elevati, fino a 45 mg/l, soprattutto nei comuni di San Martino di Lupari, Tombolo e Cittadella. L'inquinamento da composti ha origine in sversamenti puntuali, si sono notati infatti valori di concentrazione più elevata circondati da valori più bassi. Non è risultato di particolare interesse il confronto tra insiemi di dati riguardanti anni diversi e differenti profondità, poiché elude dallo scopo della tesi.

# Capitolo 1

## Le acque sotterranee

Parte dell'acqua piovana e parte dell'acqua trasportata dai fiumi si infiltra nel sottosuolo, penetra in profondità, fino a che raggiunge un livello di suolo impermeabile, andando a costituire degli accumuli di acqua sotterranea, la falda, contenuti negli spazi vuoti (pori e fratture) del sottosuolo.

Nella Pianura Padana le falde sono contenute in materiali sciolti, come sabbie e ghiaie, che costituiscono le parti del sottosuolo, dette acquiferi, in cui può accumularsi acqua sotterranea.

La pianura veneta è principalmente il risultato dell'erosione di materiali provenienti dalla montagna, del loro trasporto e della loro deposizione, a opera dei corsi d'acqua. Altre sedimentazioni sono avvenute in ambiente palustre, lagunare e marino in tempi geologicamente recenti.

Nello studio delle acque sotterranee un parametro fondamentale è la permeabilità che si misura in metri al secondo e ci indica la facilità con cui i terreni si lasciano attraversare dall'acqua. La permeabilità di solito è tanto maggiore quanto maggiori sono le dimensioni dei granuli e quanto più essi si presentano uniformi. I terreni argillosi sono praticamente impermeabili mentre quelli sabbiosi e ghiaiosi sono permeabili. Quindi le falde nel sottosuolo si trovano laddove vi sono ciottoli, ghiaie e sabbie.

Le falde possono essere in pressione oppure a pelo libero e di seguito ne evidenziamo le caratteristiche.

**FALDA FREATICA A PELO LIBERO** In questo caso la falda è alloggiata in terreni permeabili (sabbiosi-ghiaiosi-ciottolosi) e non ha un tetto di terreni impermeabili (argillosi), che impediscano le oscillazioni del livello dell'acqua nel

tempo. La sua superficie è quindi libera di muoversi e tende a salire ogni volta che viene alimentata da nuove quantità d'acqua.

Se si misura la profondità dell'acqua in un pozzo, si ottiene la profondità a cui si trova effettivamente l'acqua nel sottosuolo.

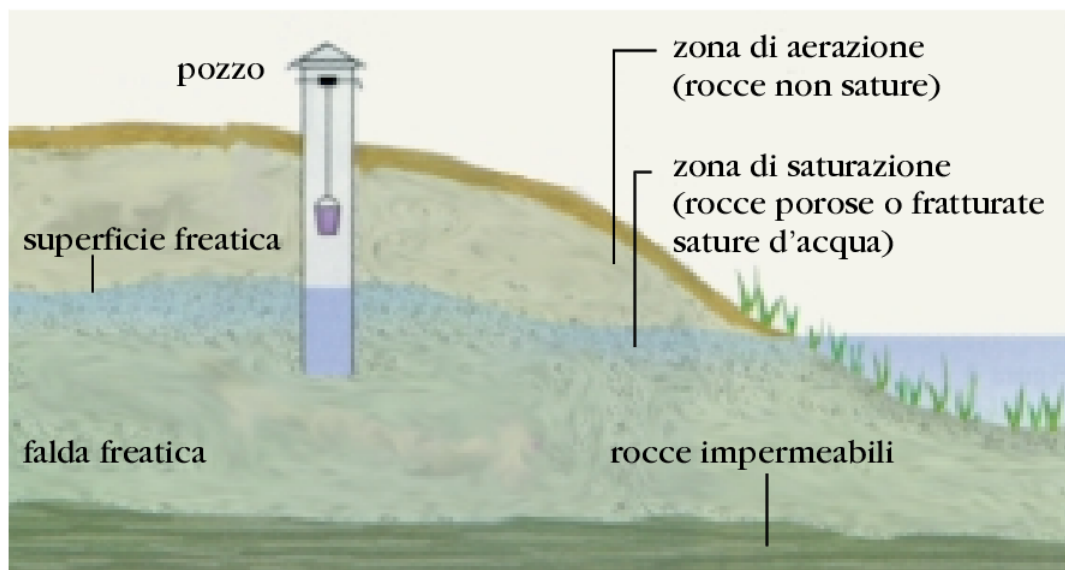


Figura 1.1: Schema di falda freatica a pelo libero

**FALDA IN PRESSIONE** Diverso è il caso che si presenta quando la falda è in pressione. Ciò si verifica quando essa, pur sempre alloggiata in terreni permeabili, è sormontata da uno strato impermeabile. Qualora questo impedisca il libero innalzarsi della superficie della falda, essa resta strettamente confinata nello strato acquifero ed esercita delle pressioni (sottopressioni), dal basso verso l'alto, sullo strato impermeabile.

Se nel corso della realizzazione di un pozzo idrico, si perfora questo strato impermeabile raggiungendo l'acquifero, la falda in tale punto tenderà a salire lungo il pozzo fino a raggiungere un livello di equilibrio detto equilibrio piezometrico. Nel caso il livello piezometrico si trovi al di sopra del piano campagna, il pozzo, una volta creato, erogherà acqua senza bisogno di installare una pompa.

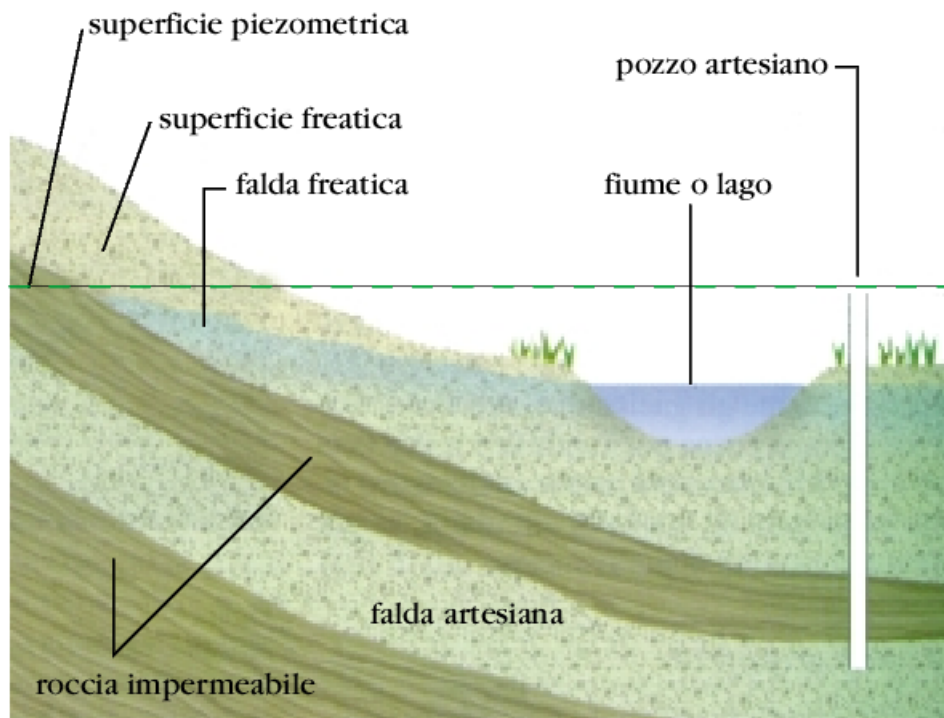


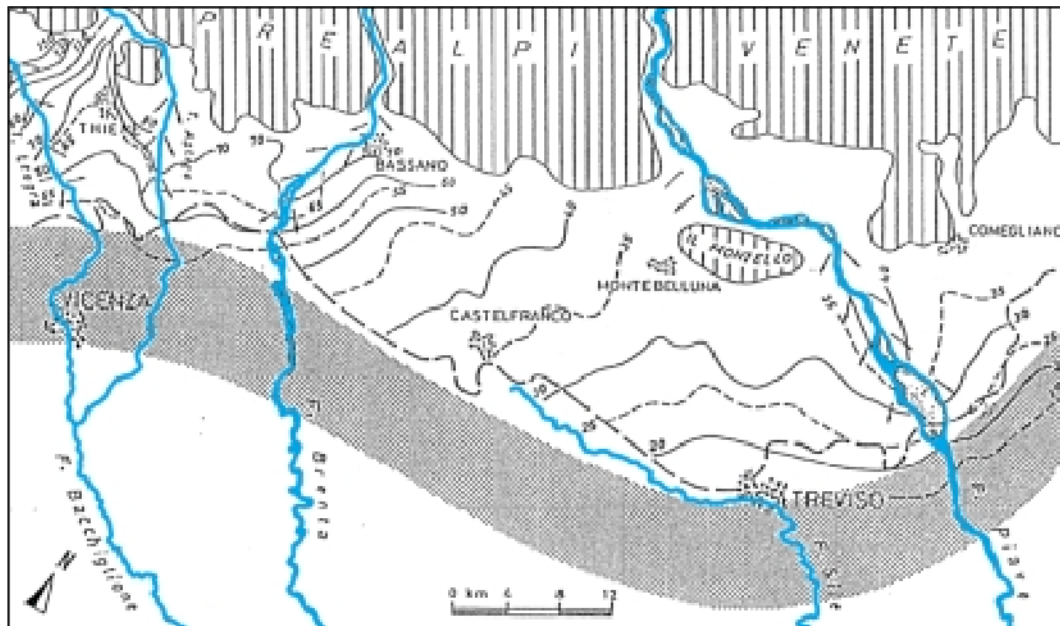
Figura 1.2: Schema di falda freatica in pressione

## 1.1 Le acque sotterranee nella pianura veneta

La pianura veneta è costituita da sedimenti quali ciottoli, ghiaie, sabbie, limi, argille e torbe di diversa origine, che hanno colmato il golfo che esisteva un tempo al suo posto. L'origine di questi sedimenti è soprattutto fluviale e fluvio-glaciale. La pianura è quindi il risultato del processo di erosione, trasporto e sedimentazione dei materiali erosi dalla catena alpina.

L'assetto idrogeologico dell'intera pianura veneta può essere schematizzato dividendo la pianura in tre fasce:

1. alta pianura;
2. media pianura;
3. bassa pianura.






-  fascia della media pianura (falde in pressione)
-  linee isofreatiche
-  tronco d'alveo disperdente

Figura 1.3: Schema idrogeologico della pianura veneta

**L'ALTA PIANURA** Nella prima fascia alla base delle Prealpi, che corrisponde all'alta pianura, si sono depositati i materiali più pesanti e grossolani, caratterizzati da elevata permeabilità.

L'alta pianura contiene una falda freatica con superficie del pelo d'acqua libera, posta a una profondità che decresce andando dal piede dei rilievi verso valle.

L'alimentazione della falda deriva soprattutto dalle dispersioni dei fiumi (Adige, Astico, Leogra, Brenta, Piave): parte della loro acqua si infiltra negli spazi tra le ghiaie e le sabbie e penetra in profondità, fino ad arrivare, appunto, alla falda.

**LA MEDIA PIANURA** Nella zona della media pianura si alternano sedimenti ghiaiosi-sabbiosi e materiali più fini. Vi è una progressiva diminuzione delle ghiaie, suddivise in diversi strati, separati da livelli di materiali fini limoso-



argillosi, a permeabilità bassissima. Di conseguenza la falda presente nell'alta pianura si suddivide in più falde sovrapposte, la prima libera e quelle sottostanti in pressione. Il sistema delle falde in pressione è collegato, verso monte, alla falda freatica che lo alimenta e ne condiziona la qualità.

La media pianura è caratterizzata dall'esistenza di pozzi artesiani, attraverso i quali l'acqua fuoriesce per pressione naturale, senza l'ausilio di pompe. In questa zona la superficie freatica, nel suo progressivo innalzamento verso il piano campagna, emerge in superficie, creando una fascia di caratteristiche sorgenti di pianura, dette risorgive.

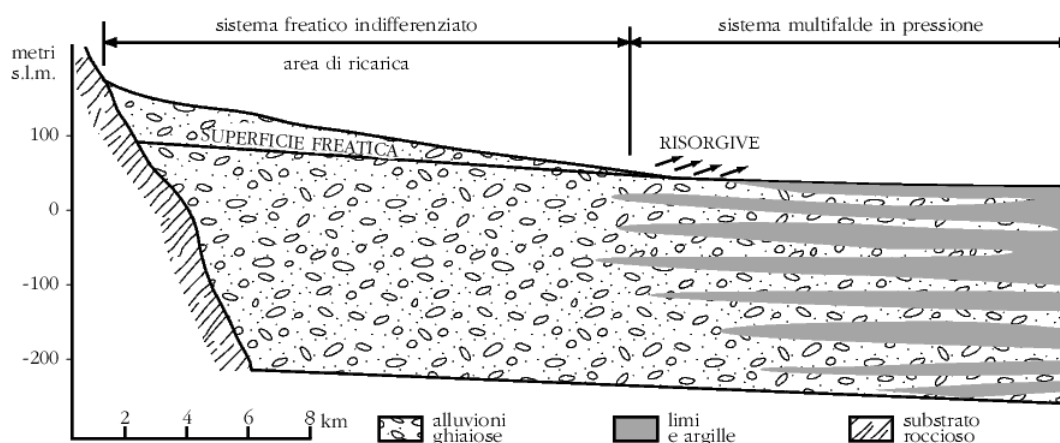


Figura 1.4: Modello idrogeologico della pianura veneta. (Sezione in direzione nord-sud)

**LA BASSA PIANURA** Nelle aree più lontane dalle Prealpi, quelle della bassa pianura, il sottosuolo è costituito in prevalenza da materiali fini, come argilla e limi, di scarsa permeabilità, entro cui si intercalano livelli sabbiosi (legati ad esempio a paleovalvei e a dune sepolte).

I corpi acquiferi presenti sono in genere costituiti da falde in sabbia, non molto estese nel sottosuolo. L'alimentazione delle falde è attribuibile principalmente a tre fattori: la dispersione dei corsi d'acqua, l'infiltrazione delle piogge, l'infiltrazione a valle delle acque di ruscellamento di versante, cioè quelle acque che scorrono lungo i versanti delle montagne e, una volta giunte a valle, si infiltrano nel sottosuolo.

Per quanto riguarda gli esigui corpi acquiferi liberi superficiali, la loro alimentazione è affidata ad apporti esclusivamente locali, quali gli afflussi meteorici e la dispersione di acqua utilizzata per irrigazioni.

## 1.2 Inquinanti delle acque sotterranee

Le acque sotterranee sono più difficilmente inquinabili di quelle superficiali, infatti possiedono la protezione naturale data dai terreni sovrastanti la falda. Una volta inquinate, però, la contaminazione persiste per tempi molto lunghi, anche decine di anni. Esistono due tipi di inquinamento:

- microbiologico,
- chimico.

L'inquinamento microbiologico è dovuto alla presenza di batteri, virus e protozoi, alcuni dei quali sono causa di malattie, anche gravi, per l'uomo. In genere, è assente nelle acque sotterranee dato che, durante il loro cammino nel sottosuolo, esse subiscono una filtrazione naturale che elimina eventuali microrganismi.

Inquinamenti microbiologici sono possibili nelle falde freatiche poco profonde e quindi più esposte alla possibile infiltrazione di inquinanti dalla superficie.

Diverso è il caso dell'inquinamento di tipo chimico. Le sostanze chimiche esistenti, potenzialmente in grado di inquinare le acque, sono moltissime. Le principali sostanze inquinanti, che si trovano con maggiore frequenza nelle acque sotterranee sono i nitrati, i nitriti, i solfati, i cloruri e i microinquinanti, vale a dire delle sostanze presenti in piccolissime concentrazioni. I più importanti microinquinanti sono: i metalli pesanti (cadmio, cromo, mercurio, piombo), gli erbicidi, i solventi organo-alogenati e altri composti organici quali il benzene, il cloruro di vinile e gli idrocarburi.

## 1.3 Sorgenti di inquinamento

Le sorgenti di inquinamento si possono dividere in due categorie:

- sorgenti di tipo puntuale,

- sorgenti di tipo diffuso.

Le sorgenti di tipo puntuale sono punti specifici in cui ha origine un inquinamento chimico, ad esempio uno scarico industriale o delle sostanze provenienti da una discarica. Gli inquinanti lentamente si infiltrano nel sottosuolo fino a raggiungere le acque della falda, dove si spostano con il lento movimento delle acque sotterranee.

Sorgenti di questo tipo producono inquinamenti localizzati su aree ristrette, ma con elevate concentrazioni. I più diffusi inquinanti delle acque sotterranee da fonti puntuali risultano essere i solventi organo-alogenati (tricloroetilene, tetracloroetilene e tricloroetano) e il cromo. L'origine di questi inquinamenti è da ricercare in sversamenti, discariche abusive, perdite accidentali durante i cicli di lavorazione, legate per lo più a piccole e medie industrie.

Le sorgenti di tipo diffuso sono tipicamente legate all'attività agricola. Devono il loro nome al fatto che non è presente un punto specifico di inquinamento, ma le sostanze tossiche provengono da tutta l'area coltivata. Infatti, parte dei fertilizzanti (nitrati), dei pesticidi e degli erbicidi sparsi sul terreno può essere trasportata fino alla falda dall'acqua che si infiltra nel sottosuolo. Questo tipo di inquinamento comporta in genere concentrazioni basse su aree vaste.

## 1.4 Movimento di un inquinante nel sottosuolo

Un inquinante abbandonato su un terreno si infiltra nel sottosuolo dove possono avvenire una serie di processi che fermano, ritardano o attenuano l'inquinamento. Il suolo, infatti, possiede una capacità di autodepurazione: attraverso processi chimici, fisici e biologici gli inquinanti vengono parzialmente bloccati, degradati o resi non tossici.

Una volta arrivato in falda, però, l'inquinante vi rimane e difficilmente intervengono reazioni chimiche che possano eliminarlo. Di conseguenza inizia a muoversi con le acque sotterranee seguendo la direzione dell'acqua della falda. Le velocità di propagazione delle sostanze inquinanti dipendono dalle velocità di deflusso delle acque sotterranee, che nell'alta pianura veneta sono di alcuni metri al giorno, mentre risultano lentissime, pochi centimetri al giorno, nella media pianura, ma subisce anche dei ritardi, dovuti a processi di assorbimento nel materiale poroso, e di dispersione e diffusione nelle direzioni perpendicolari al flusso.

In genere la forma del pennacchio inquinante dipende dalla struttura idrogeologica, dal meccanismo di immissione dell'inquinante nel sottosuolo e dalle caratteristiche chimico-fisiche dell'inquinante.

(Per maggiori informazioni si veda [11] e [7]).

# Capitolo 2

## Modellazione geostatistica

### 2.1 Introduzione

Molti dei fenomeni naturali studiati sono variabili sia nel tempo che nello spazio e possono essere descritti mediante modelli deterministici, ma la maggior parte di essi è caratterizzata dalla presenza di variabili che interagiscono in maniera molto complessa e inducono a pensare che certi comportamenti siano casuali. Le condizioni in cui avvengono questi processi non sono sempre totalmente conosciuti e non è quindi possibile descriverli esclusivamente attraverso leggi fisiche e chimiche.

Diventa necessario introdurre metodi e modelli probabilistici e statistici, per studiare parametri naturali campionati o conosciuti solo in parte. La teoria dei processi stocastici riconosce la componente di incertezza e l'approccio probabilistico è in grado, non solo di affrontare il problema della previsione dei valori incogniti, ma anche di determinare l'accuratezza di tale previsione.

I fenomeni suddetti possono essere descritti utilizzando il concetto di variabile aleatoria. La totalità delle variabili aleatorie costituisce un processo stocastico: se si utilizza lo spazio matematico ad una sola dimensione e la variabile dipendente è il tempo, allora si tratterà di un processo stocastico temporale, se invece si utilizzano due o tre dimensioni, si parlerà di processo stocastico spaziale.

L'analisi delle serie storiche è uno dei primi campi in cui la variabilità è stata considerata e analizzata con metodi stocastici, in seguito applicati e sviluppati anche per la variabilità spaziale, dando origine alla Geostatistica. Con questo termine si indicano tutti i metodi statistici applicati alle scienze naturali e, in

particolare, quelli relativi alle variabili regionalizzate.

La teoria delle variabili regionalizzate è nata negli anni '50 del secolo scorso quando, in Sud Africa, D. Krige cominciò ad applicare tecniche statistiche per la previsione delle riserve minerarie. A partire dagli anni '60 il matematico francese G. Matheron formulò le basi teoriche di tali metodi e un modello che tenesse conto delle fluttuazioni delle variabili osservate ed, in particolar modo, della correlazione spaziale.

L' applicazione più diffusa è la previsione del valore di una grandezza in punti in cui non è possibile effettuare misurazioni. Tale obiettivo si può raggiungere utilizzando le osservazioni disponibili nei punti campionati, supponendo l'esistenza di una relazione tra valori osservati e incogniti. Proprio per questo la Geostatistica è stata inizialmente utilizzata dall'industria mineraria per ridurre gli elevati costi delle perforazioni. In seguito si sono rivelati i molteplici campi di applicazione di tale disciplina: scienze del suolo, idrogeologia, meteorologia, scienze ambientali, agricoltura, ingegneria civile, ecc.

## 2.2 Le variabili regionalizzate

Una variabile regionalizzata è la realizzazione di un processo stocastico, cioè di un insieme di variabili casuali corrispondenti ai punti di un dominio  $D$  oggetto di studio. Dunque, per ogni punto  $u$  in uno spazio  $d$  dimensionale, il valore del parametro di interesse  $z(u)$  è una delle realizzazioni del processo  $Z(u)$ .

Matheron (1971) ha definito le variabili spaziali come variabili regionalizzate per descrivere la loro natura, espressa da due aspetti complementari: uno casuale, che tiene conto di irregolarità locali presentate dal fenomeno (*piccola scala*), e un aspetto strutturale, che riflette le caratteristiche globali dello stesso (*larga scala*).

Il modello fondamentale per una variabile regionalizzata  $Z(u)$ , osservata nella localizzazione  $u$ , è:

$$Z(u) = m(u) + Y(u). \quad (2.1)$$

La componente casuale è descritta da  $Y(u)$  e quella strutturale da  $m(u)$ , chiamata trend. Nel caso in cui la componente casuale fosse assente, il fenomeno potrebbe essere descritto dal solo trend, rappresentato tramite una funzione.

Esistono anche situazioni in cui il trend  $m(u)$  è completamente assente op-

pure costante nel dominio in esame. Nonostante la presenza del trend è necessario esaminare l'informazione contenuta nella componente residua  $Y(u)$ , poiché le fluttuazioni descritte da tale componente non sono errori, ma caratteristiche fondamentali dei dati osservati. La correlazione spaziale viene studiata tenendo conto della componente residua  $Y(u)$ .

Uno degli obiettivi della Geostatistica è quello di identificare le due componenti suddette per risolvere il problema della previsione di una variabile regionalizzata, a partire dall'informazione frammentaria contenuta nel campione osservato. L'interpretazione dei parametri supporta l'ipotesi che non sia possibile descrivere completamente i fenomeni naturali attraverso regole deterministiche. Nella maggior parte dei casi è impossibile controllare l'assunto che il parametro sia la realizzazione di una funzione casuale avendo a che fare con una singola realizzazione.

Per ogni insieme di punti  $u_1, \dots, u_n$  nel dominio  $D$ , è assegnata la funzione di ripartizione  $F_{u_1, \dots, u_n}$ . Utilizzando queste funzioni per tutti gli insiemi di possibili valori  $w_1, \dots, w_n$  si può trovare la probabilità  $P$ :

$$P(Z(u_1) \leq w_1, \dots, Z(u_n) \leq w_n) = F_{u_1, \dots, u_n}(w_1, \dots, w_n). \quad (2.2)$$

Nel dominio  $D$  c'è un elevato numero di sottoinsiemi finiti, e poiché per ciascun punto del dominio solitamente è disponibile solo un valore (la singola realizzazione, il valore campionato), la previsione della funzione di ripartizione, basata sui dati campionati, non sembra realistica. Anche nel caso di variabili con misure ripetute (ad esempio per la qualità delle acque sotterranee), esse non sono comunque sufficienti.

E' necessario allora imporre delle ipotesi di stazionarietà, tramite le quali è possibile ritenere il fenomeno omogeneo, permettendo l'inferenza statistica.

Una prima ipotesi che si può introdurre è la stazionarietà forte. La funzione casuale  $Z(u)$  è stazionaria in senso forte se per ogni insieme di punti  $u_1, \dots, u_n$  del dominio  $D$ , per ogni insieme di possibili valori  $w_1, \dots, w_n$  e per ogni vettore  $h$  si ha:

$$\begin{aligned} P(Z(u_1) \leq w_1, \dots, Z(u_n) \leq w_n) = \\ = P(Z(u_1+h) \leq w_1, \dots, Z(u_n+h) \leq w_n). \end{aligned} \quad (2.3)$$

Questa equazione significa che la distribuzione della funzione casuale non dipende dalla localizzazione, dunque valori rilevati in punti distinti possono essere considerati come realizzazioni della stessa variabile regionalizzata.

Altre ipotesi più deboli di stazionarietà, legate ai momenti del processo, sono:

- la stazionarietà del secondo ordine;
- la stazionarietà intrinseca.

La stazionarietà del secondo ordine assume :

$$E[Z(u)] = m, \quad (2.4)$$

per ogni  $u \in D$ , cioè il valore atteso della funzione casuale  $Z(u)$  è costante in tutto il dominio  $D$ , e

$$E[(Z(u+h) - m)(Z(u) - m)] = C(h), \quad (2.5)$$

per ogni  $u, u+h \in D$ , cioè la funzione di covarianza  $C(h)$  di due variabili casuali (detta *covariogramma*), corrispondente a due siti diversi  $u$  e  $u+h$ , dipende solo dalla distanza  $|h|$  che separa i due punti. Quando  $h = 0$  si ha:

$$C(0) = E[(Z(u) - m)(Z(u) - m)] = Var[Z(u)]. \quad (2.6)$$

Quindi le variabili casuali corrispondenti a punti diversi del dominio non hanno solo lo stesso valore atteso, ma devono avere anche la stessa varianza finita.

Raramente i fenomeni fisici hanno queste caratteristiche, ma è possibile formulare un'ipotesi più debole: la stazionarietà intrinseca. Questa comprende la (2.4) e inoltre :

$$\frac{1}{2}Var[Z(u+h) - Z(u)] = \frac{1}{2}E[(Z(u+h) - Z(u))^2] = \gamma(h), \quad (2.7)$$

per ogni  $u, u+h \in D$ , dipende solo dalla distanza  $|h|$  che separa i due punti, e  $[Z(u+h) - Z(u)]$  ha varianza finita indipendente da  $u$ . La funzione  $\gamma(h)$  è denominata *semivariogramma* e in geostatistica descrive la correlazione spaziale.

Si può fare riferimento equivalentemente a variogramma o semivariogramma, poiché si differenziano esclusivamente per un fattore 2 di proporzionalità.



E' importante notare che la stazionarietà del secondo ordine implica la stazionarietà intrinseca, ma non vale il viceversa. Quando è verificata la stazionarietà del secondo ordine si ha:

$$\begin{aligned} E[(Z(u+h) - Z(u))^2] &= \\ &= 2C(0) - 2C(h) \end{aligned} \quad (2.8)$$

e da qui la relazione:

$$\gamma(h) = C(0) - C(h). \quad (2.9)$$

## 2.3 Proprietà del variogramma

Il grafico di  $\gamma(h)$  generalmente, quando valgono le ipotesi di stazionarietà del secondo ordine, cresce dal valore di origine  $\gamma(0) = 0$ , aumenta con il crescere della distanza  $|h|$  e tende a  $C(0)$ , poiché al tendere di  $h \rightarrow \infty$ , il covariogramma è prossimo allo zero. Infatti se  $Z(u)$  e  $Z(u+h)$  sono indipendenti, allora  $C(h) = 0$  e per la (2.9) si ha

$$\gamma(h) = C(0), \quad (2.10)$$

per  $|h| \geq a$ , dove  $a$  viene denominato *range*, cioè la distanza oltre la quale è possibile considerare trascurabile il covariogramma e il variogramma assume valori prossimi alla varianza del fenomeno. In altre parole il *range* rappresenta la massima distanza entro la quale si manifesta correlazione tra semivarianza e *lag*  $h$ , cioè l'autocorrelazione spaziale della variabile considerata. In questo caso il variogramma raggiunge il valore limite, chiamato *sill*.

Se non è verificata la stazionarietà del secondo ordine, ma vale la stazionarietà intrinseca, si hanno variogrammi senza *sill*.

Il variogramma può essere descritto dai parametri: *range*, *sill* e *nugget*. Il *nugget* costituisce la parte non spiegata della semivarianza, imputabile alla variabilità casuale (errori di laboratorio, errori strumentali,...) e alla variabilità spaziale presente a distanza minore di quella dei siti campionati. Se per  $h \rightarrow 0$  il variogramma tende ad un valore  $c > 0$  allora  $c$  è detto *nugget*.

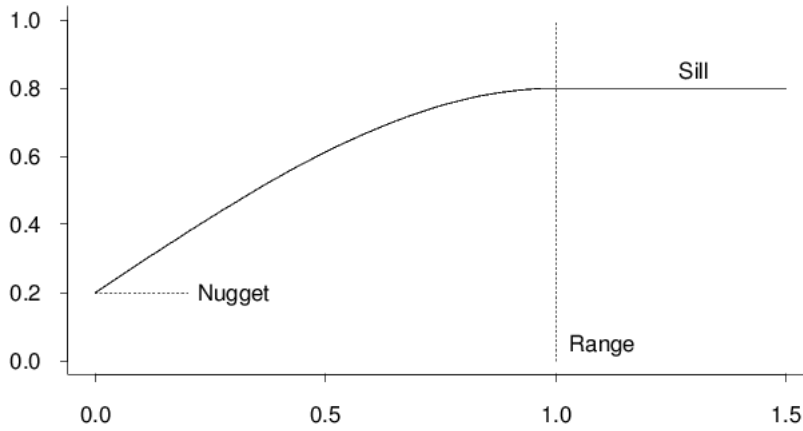


Figura 2.1: Forma idealizzata di variogramma con *nugget*, *range* e *sill*

Per ogni combinazione lineare  $\sum_{i=1}^n \theta_i Z(u_i)$ , tale che  $\sum_{i=1}^n \theta_i = 0$ , la varianza è finita e può essere calcolata come:

$$\text{Var} \left[ \sum_{i=1}^n \theta_i Z(u_i) \right] = - \sum_{j=1}^n \sum_{i=1}^n \theta_j \theta_i \gamma(u_i - u_j). \quad (2.11)$$

Poiché la varianza non può essere negativa, l'equazione (2.11) dà una condizione necessaria per il variogramma, cioè per ogni peso  $\theta_i$ , per cui  $\sum_{i=1}^n \theta_i = 0$ , allora  $-\gamma(h)$  deve essere una funzione definita positiva in modo condizionato:

$$- \sum_{j=1}^n \sum_{i=1}^n \theta_j \theta_i \gamma(u_i - u_j) \geq 0. \quad (2.12)$$

Si può provare che questa condizione è solo sufficiente. La disuguaglianza (2.12) può essere verificata solo per un numero finito di combinazioni di  $u_i$  e  $\theta_i$ .

Per questa proprietà, il semivariogramma, quando  $h \rightarrow \infty$ , deve crescere più lentamente di  $|h|^2$ , se invece si comporta allo stesso modo non possono valere le ipotesi della stazionarietà intrinseca. Un comportamento di questo tipo indica la presenza di un *trend*, lungo una certa direzione definita dal vettore  $h$ .

### 2.3.1 Isotropia e anisotropia

Si parla di isotropia se il variogramma dipende solo dalla lunghezza del vettore  $h$  e non varia con la direzione spaziale. In tal caso il variogramma sperimentale

può essere calcolato sotto la sola condizione  $|u_i - u_j| = |h|$ .

Si può osservare se la variabile regionalizzata ha continuità spaziale che dipende dalla direzione calcolando e confrontando i variogrammi empirici corrispondenti a diverse direzioni  $\alpha$  dello spazio, stabilendo un angolo di tolleranza  $\Delta\alpha$ . Se si notano direzioni preferenziali, lungo le quali il fenomeno si sviluppa con variabilità più accentuata delle altre, allora la variabile regionalizzata è detta anisotropa. Le anisotropie possono essere classificate in due categorie:

- anisotropia geometrica;
- anisotropia zonale.

### 2.3.1.1 Anisotropia geometrica

La variabile regionalizzata che presenta anisotropia geometrica può essere resa isotropa tramite una trasformazione lineare  $T$  delle coordinate, tale per cui  $Z(u') = Z(Tu)$  diventa isotropa.

Dunque una semplice trasformazione delle coordinate conduce ad una situazione in cui ha un ruolo solo la distanza (nel nuovo sistema di coordinate). L'esistenza di tale trasformazione implica che il valore di *sill*, se esiste, sia lo stesso in tutte le direzioni.

Disegnando i *range* corrispondenti a diverse direzioni, se essi formano un'ellisse (nel caso bidimensionale), allora una rotazione e la conseguente contrazione saranno l'appropriata trasformazione  $T$ . La trasformazione è:

$$\begin{aligned} x' &= \lambda (x \cos \varphi + y \sin \varphi) \\ y' &= -x \sin \varphi + y \cos \varphi, \end{aligned} \tag{2.13}$$

dove  $\varphi$  è l'angolo formato dall'asse  $x$  e l'asse principale dell'anisotropia (ellisse), mentre  $\lambda$  è il rapporto tra i due *range* ortogonali, che rappresentano la più alta e la più bassa variabilità.

Nel caso tridimensionale l'ellisse è sostituita da un ellissoide. La variabilità nella direzione verticale è più elevata rispetto a quella orizzontale, e ciò conduce ad una forte anisotropia.

### 2.3.1.2 Anisotropia zonale

Se i *range* non formano un'ellisse, oppure i valori di *sill* sono diversi, allora si ha anisotropia zonale. In questo caso deve essere adattato un variogramma teorico annidato, in cui ogni componente ha la sua forma di anisotropia:

$$\gamma(h) = \sum_i \gamma_i(h).$$

### 2.3.2 I modelli teorici

Per confrontare variogrammi empirici con funzioni adatte, sono stati sviluppati diversi modelli teorici isotropi, che possono essere suddivisi in due gruppi a seconda che sia verificata (modelli con *sill*) o no (modelli privi di *sill*) la stazionarietà del secondo ordine.

#### 2.3.2.1 Modelli con *sill*

Si presentano i cinque tipi di variogramma con *sill* più ricorrenti (si noti che combinazioni lineari con coefficienti positivi di questi modelli sono ancora variogrammi con *sill*).

**Puro effetto *nugget*** Il puro effetto *nugget* corrisponde al caso in cui manca totalmente la correlazione spaziale tra le variabili casuali corrispondenti a diverse locazioni. Questo significa che:

$$\gamma(h) = \begin{cases} 0 & \text{se } |h| = 0 \\ C & \text{se } |h| > 0 \end{cases}, \quad (2.14)$$

dove  $C$  è il *sill*.

**Modello sferico** Questo è il modello di variogramma più utilizzato. Ha un comportamento lineare in prossimità dell'origine e presenta *range*.

La formula del variogramma sferico è:

$$\gamma(h) = \begin{cases} C \left( \frac{3}{2} \frac{r}{a} - \frac{1}{2} \left( \frac{r}{a} \right)^3 \right) & \text{se } r \in [0, a] \\ C & \text{se } r > a \end{cases}, \quad (2.15)$$

dove  $r = |h|$  e  $a$  è il valore *range*.

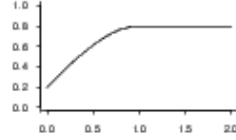


Figura 2.2: Modello di variogramma sferico

**Modello esponenziale** Come il variogramma sferico anche quello esponenziale presenta un comportamento lineare in prossimità dell'origine, ma raggiunge il valore *sill* solo asintoticamente, per cui non ammette *range*. La formula del variogramma esponenziale è:

$$\gamma(h) = C \left( 1 - \exp \left( -\frac{r}{a} \right) \right). \quad (2.16)$$

Comunque si può definire un *range* effettivo  $a' = 3a$ , per il quale  $\gamma(a') = 0,95C$ .

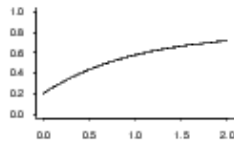


Figura 2.3: Modello di variogramma esponenziale

**Modello gaussiano** Anche questo modello raggiunge asintoticamente il valore di *sill*. Come quello esponenziale, il variogramma gaussiano non prevede un limite teorico tra le variabili casuali correlate e quelle non correlate. Assume un comportamento parabolico in prossimità dell'origine e descrive un fenomeno molto regolare. La formula del variogramma gaussiano è:

$$\gamma(h) = C \left( 1 - \exp \left( - \left( \frac{r}{a} \right)^2 \right) \right). \quad (2.17)$$

Il *range* effettivo in questo caso è  $a' = \sqrt{3}a$ , per il quale  $\gamma(a') = 0,95C$ .

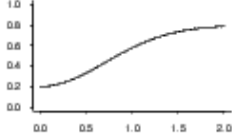


Figura 2.4: Modello di variogramma gaussiano

**Modello *wave*** Questo modello è l'unico non monotono degli esempi. Può essere rappresentato dalla formula:

$$\gamma(h) = C \left( 1 - \frac{a}{r} \sin \left( \frac{r}{a} \right) \right). \quad (2.18)$$

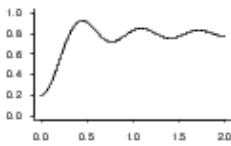


Figura 2.5: Modello di variogramma *wave*

### 2.3.2.2 Modelli privi di *sill*

Quando una variabile regionalizzata non verifica la stazionarietà del secondo ordine, ma solo le ipotesi intrinseche, il variogramma può presentare un incremento illimitato. Questi modelli non ammettono né varianza, né covarianza.

**Funzione potenza** Un valido modello di variogramma è rappresentato dalla funzione:

$$\gamma(h) = w |h|^\lambda \quad \text{con } 0 < \lambda < 2. \quad (2.19)$$

In figura è il modello con  $\lambda = 0.5$ . Per  $\lambda = 1$  si ha il variogramma lineare, ed è usato abbastanza spesso in geostatistica.



Figura 2.6: Modelli di variogramma potenza e lineare

**Modelli complessi** Tutti i modelli visti finora soddisfano l'equazione (2.12), ma non sempre si possono utilizzare per descrivere la variabile regionalizzata oggetto di studio. Per arricchire l'insieme dei variogrammi teorici si possono combinare i precedenti modelli in strutture annidate (*nested*). Tramite queste strutture, entrano in gioco simultaneamente tutte le cause che influiscono sulla variabilità spaziale di un fenomeno a distanze diverse.

I modelli di questo tipo sono rappresentati da una somma di variogrammi:

$$\gamma(h) = \gamma_0(h) + \gamma_1(h) + \dots + \gamma_i(h).$$

## 2.4 Il variogramma empirico

Il variogramma empirico è una stima del variogramma teorico definito nell'equazione (2.7) e può essere calcolato per un numero finito di vettori  $h$ . Rappresenta la relazione tra variabilità spaziale (*semivarianza*) e  $h$ . La semivarianza è la varianza dell'incremento quando le località di osservazione, prese a due a due, sono separate da una distanza vettoriale detta *lag*.

Il variogramma deve essere stimato sulla base dei dati disponibili per l'analisi, allo scopo di indagare le caratteristiche del fenomeno regionalizzato. Nel caso si disponga di dati su una griglia regolare, il variogramma empirico può essere stimato con:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{u_i - u_j = h} (Z(u_i) - Z(u_j))^2, \quad (2.20)$$

dove  $N(h)$  è il numero di coppie di siti separati dal vettore  $h$ <sup>1</sup>. Quando i siti sono disposti in modo irregolare è necessario modificare la definizione di  $N(h)$ :

$$N(h) \equiv \{(u_i, u_j) : u_i - u_j \in T(h); i, j = 1, \dots, n\},$$

dove  $T(h)$  rappresenta una regione di tolleranza intorno ad  $h$ .

La stima del variogramma empirico è molto sensibile a valori estremi, poiché la distribuzione dei quadrati delle differenze è molto asimmetrica. In molte applicazioni (ad esempio quando si analizzano dati di tipo ambientale) in cui la distribuzione dei valori è caratterizzata da un coefficiente di asimmetria sufficientemente elevato, il semivariogramma non è la misura più appropriata per studiare la correlazione spaziale.

E' noto come in tali casi la media aritmetica non sia il miglior stimatore, ma possa essere sostituita da altri più robusti. Cressie e Hawkins (1980) suggeriscono:

$$\hat{\gamma}^*(h) = \frac{1}{2} \left( \frac{1}{N(h)} \sum_{(i,j) \in R(h)} \sqrt{|Z(u_i) - Z(u_j)|} \right)^4 (0,457 + \frac{0,494}{N(h)})^{-1}. \quad (2.21)$$

Questo stimatore risulta efficiente se la distribuzione dei dati grezzi è simile ad una Normale nella parte centrale e si discosta da essa nelle code.

### 2.4.1 Nuvola del variogramma

L'insieme delle distanze calcolate per tutte le coppie di dati combinate con le differenze al quadrato delle osservazioni

$$\{(u_i - u_j, (Z(u_i) - Z(u_j))^2, i \neq j; i, j = 1, \dots, n\}$$

---

<sup>1</sup>30 coppie è un numero ragionevole per ottenere una buona stima [5].



possono essere rappresentate mediante un grafico, detto *nuvola del variogramma*. Viene rappresentata come il diagramma di dispersione delle coppie

$$\{|u_i - u_j|, (Z(u_i) - Z(u_j))^2, i \neq j; i, j = 1, \dots, n\}.$$

Grazie a questo grafico si possono individuare alcune caratteristiche importanti:

- valori anomali globali: misurazioni separate dalla parte principale dei dati. Si notano nella nuvola poiché, per ogni distanza, le differenze al quadrato delle coppie che individuano il valore anomalo sono significativamente più lontane da tutte le altre;
- valori anomali locali: sono celati nella parte principale dei dati osservati, ma si distinguono dai valori vicini. Si trovano vicino all'origine con elevate differenze al quadrato, mentre a distanze medio-lunghe hanno un andamento normale;
- non stazionarietà locale: si possono individuare piccole aree nel dominio in cui la stazionarietà non è verificata. Si può notare se c'è un insieme di punti che possiede maggiore variabilità rispetto ai punti circostanti;
- assenza di distribuzione normale: indica la presenza di pochi valori per cui la variabile osservata si può presumere sia discreta. La nuvola del variogramma presenta righe orizzontali laddove non ci sono valori distinti sufficienti per tutte le differenze.

## 2.5 Aspetti inferenziali

Una volta scelto un modello teorico di variogramma, i metodi di stima dei parametri che possono essere utilizzati sono:

- la massima verosimiglianza;
- i minimi quadrati;
- i minimi quadrati generalizzati.

Il metodo maggiormente utilizzato è quello dei minimi quadrati generalizzati poiché si ottengono stimatori consistenti e asintoticamente normali, consentendo di generare intervalli di confidenza.

**MASSIMA VEROSIMIGLIANZA** Si assume

$$Z = (Z(u_1), \dots, Z(u_n))' \sim N(X\beta, \Sigma(\theta))$$

con rango  $q < n$ . Le stime di massima verosimiglianza di  $\beta$  e  $\theta$  si ricavano minimizzando la funzione di verosimiglianza

$$L(\beta, \theta) = \frac{n}{2} \log(2\pi) + \frac{1}{2} \log |\Sigma(\theta)| + \frac{1}{2} (Z - X\beta)' (Z - X\beta),$$

quindi

$$L(\hat{\beta}, \hat{\theta}) = \inf \{L(\beta, \theta) : \beta \in R^q, \theta \in \Theta\}.$$

**MINIMI QUADRATI** Si ottiene una stima di  $\theta$  tramite

$$\hat{\theta}_{MQ} = \operatorname{argmin}_{\theta \in \Theta} \left\{ \sum_{i=1}^n (\hat{\gamma}(h_i) - \gamma(h_i; \theta))^2 \right\}.$$

**MINIMI QUADRATI GENERALIZZATI** La stima è data da

$$\hat{\theta}_{MQ} = \operatorname{argmin}_{\theta \in \Theta} \left\{ \sum_{i=1}^n |N(h)| \left( \frac{\hat{\gamma}(h_i)}{\gamma(h_i; \theta)} - 1 \right)^2 \right\}.$$

Questo metodo di stima è sensibile al numero di differenze quadratiche presenti in ogni intervallo  $h_i$  di distanze con le quali si è calcolato il variogramma empirico, infatti pesa i valori del variogramma proporzionalmente alla numerosità  $N(h)$ .

## 2.6 Il kriging

I variogrammi forniscono molte informazioni sui parametri oggetto di studio e permettono la previsione dei parametri in siti non campionati. Il *kriging* è la procedura per ottenere tali risultati. Si tratta di una media mobile pesata, che permette di effettuare previsioni non distorte attraverso un predittore lineare, ricavato per interpolazione da osservazioni localizzate nei pressi dei punti non

campionati. La conclusione dell'analisi geostatistica avviene con la restituzione cartografica dei valori dell'attributo studiato nella regione indagata, ottenuti appunto tramite il *kriging*.

## 2.7 *Kriging* ordinario

Il *kriging* è una tecnica di interpolazione con importanti proprietà:

1. Il *kriging* è un interpolatore esatto, infatti per ogni punto osservato  $u_i$  si ha  $Z(u_i) = Z^*(u_i)$  e la corrispondente varianza di previsione è uguale a zero.
2. I pesi del *kriging* sono calcolati con l'aiuto del variogramma, della localizzazione dei punti di misurazione e dei punti da prevedere. Non vengono considerate solo le distanze tra i punti osservati e quelli da prevedere, ma anche le posizioni relative dei punti.
3. La somma dei pesi del *kriging* è uguale a 1, ma possono anche essere negativi. Perciò non è vera l'ipotesi:

$$\max \{Z(u_i)\} \leq Z^*(u) \leq \min \{Z(u_i)\}.$$

4. I pesi del *kriging* non sono influenzati dai valori osservati. Se c'è la stessa configurazione in due siti diversi, i pesi saranno gli stessi, indipendentemente dai valori misurati. I valori osservati influenzano il variogramma che è la base per il calcolo dei pesi del *kriging*.
5. Punti lontani ricevono pesi di *kriging* inferiori se sono disponibili osservazioni più vicine.

Il *kriging* fornisce non solo i valori previsti, ma anche le corrispondenti varianze che esprimono l'affidabilità della previsione [6].

### 2.7.1 *Kriging* puntuale

Uno dei più comuni problemi di interpolazione è la previsione di un sito  $u$  non campionato. Bisogna ricavare un previsore lineare, cioè una combinazione lineare di valori della variabile regionalizzata in siti noti, del tipo

$$Z^*(u) = \sum_{i=1}^n \lambda_i Z(u_i). \quad (2.22)$$

I pesi  $\lambda_i$  vengono scelti in modo da avere una previsione non distorta e con la minore varianza. Per la stazionarietà del secondo ordine si ha:

$$E[Z(u)] = m \quad \text{per ogni } u \in D. \quad (2.23)$$

Dunque per il previsore lineare

$$E[Z^*(u)] = \sum_{i=1}^n \lambda_i E[Z(u_i)] = m, \quad (2.24)$$

da cui

$$\sum_{i=1}^n \lambda_i = 1 \quad (2.25)$$

è la condizione di non distorsione. Utilizzando le ipotesi della stazionarietà del secondo ordine, la varianza di previsione si può calcolare grazie alla funzione di covarianza  $C(h)$  come

$$\begin{aligned} \sigma^2(u) &= \text{Var}[Z(u) - Z^*(u)] = E \left[ \left( Z(u) - \sum_{i=1}^n \lambda_i Z(u_i) \right)^2 \right] = \\ &= E \left[ Z(u)^2 + \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j Z(u_i) Z(u_j) - 2 \sum_{i=1}^n \lambda_i Z(u_i) Z(u) \right] = \\ &= C(0) + \sum_{j=1}^n \sum_{i=1}^n \lambda_j \lambda_i C(u_i - u_j) - 2 \sum_{i=1}^n \lambda_i C(u_i - u). \end{aligned} \quad (2.26)$$

Per minimizzare una funzione di  $n$  variabili si eguagliano a zero le  $n$  derivate parziali rispetto alle  $n$  incognite, ottenendo un sistema che va risolto tenendo conto del vincolo sui pesi  $\lambda_i$  (condizione di non distorsione). Si risolve perciò un problema di minimo vincolato con il metodo dei moltiplicatori di Lagrange. Deve essere minimizzata la funzione

$$\sigma^2(u) - 2\mu \left( \sum_{i=1}^n \lambda_i - 1 \right), \quad (2.27)$$

calcolando le derivate parziali rispetto ai parametri  $\lambda_i$  e  $\mu$  ignoti ed uguagliandole a zero.

Si deve risolvere il sistema di equazioni lineari:

$$\begin{cases} \sum_{j=1}^n \lambda_j C(u_i - u_j) - \mu = C(u_i - u) & i = 1, \dots, n \\ \sum_{j=1}^n \lambda_j = 1 \end{cases}, \quad (2.28)$$

detto sistema del *kriging* stazionario, dal quale si ottengono i pesi  $\lambda_i$  per il previsore lineare. Il sistema è regolare e ammette una soluzione unica se e solo se la matrice delle covarianze è strettamente definita positiva, ossia se ha determinante diverso da zero.

Si può esprimere la varianza di previsione mediante il variogramma:

$$\begin{aligned} \sigma^2(u) &= \text{Var} [Z(u) - Z^*(u)] = \\ &= - \sum_{j=1}^n \sum_{i=1}^n \lambda_j \lambda_i \gamma(u_i - u_j) + 2 \sum_{i=1}^n \lambda_i \gamma(u_i - u). \end{aligned} \quad (2.29)$$

L'obiettivo è quello di minimizzare  $\sigma^2(u)$ , sotto le condizioni di non distorsione, e allora il sistema del *kriging* diventa:

$$\begin{cases} \sum_{j=1}^n \lambda_j \gamma(u_i - u_j) + \mu = \gamma(u_i - u) & i = 1, \dots, n \\ \sum_{j=1}^n \lambda_j = 1 \end{cases}. \quad (2.30)$$

## 2.8 Metodi non stazionari

Molte variabili naturali non si adattano alle ipotesi intrinseche, perché ci sono cambiamenti sistematici che influenzano il variogramma empirico e conducono a risultati non accettabili. Ad esempio per le acque sotterranee ci sono variabili che hanno una tendenza sistematica e perciò non possono essere studiati con metodi stazionari. Si suppone allora che non sia verificata l'equazione (2.4) della stazionarietà intrinseca e quindi, invece di un valore atteso costante, si è in presenza di un *trend*. E' possibile trattare il *trend* mediante il *kriging* universale.

### 2.8.1 Kriging universale

Nel caso in cui una variabile regionalizzata presenta un *trend* lungo una certa direzione, un modello consiste nel considerare il *trend*  $m(u)$  come una forma polinomiale del tipo:

$$m(u) = \sum_{s=0}^S b_s m_s(u), \quad (2.31)$$

dove  $m_0(u) = 1$ , i coefficienti  $b_s$  sono ignoti, e devono essere ricavati per  $s > 0$ . Si suppone che la (2.31) non valga per l'intero dominio oggetto di studio, ma solo per i siti vicini.

Per la modellazione del *trend*  $m(u)$  tipicamente vengono utilizzati polinomi di primo grado delle coordinate  $x_1$  e  $x_2$ :

$$m(x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2, \quad (2.32)$$

oppure polinomi di secondo grado:

$$m(x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 (x_1)^2 + \beta_4 (x_2)^2 + \beta_5 x_1 x_2, \quad (2.33)$$

i cui coefficienti si possono stimare con il metodo dei minimi quadrati.

La funzione aleatoria non stazionaria viene espressa come somma del *trend*  $m(u)$  e di un termine residuo. Il kriging non stazionario, come quello ordinario, richiede la conoscenza del variogramma, ma in questo caso è il variogramma dei residui  $Y(u) = Z(u) - \widehat{m}(u)$ . Dato che il *trend* non è noto, non si può stimare il variogramma direttamente dalle osservazioni.

Come nel caso stazionario bisogna prevedere un sito  $u$  non campionato. Le previsioni dei coefficienti sono combinazioni lineari dei valori misurati:

$$B_s = \sum_{i=1}^n \lambda_{i,s} Z(u_i). \quad (2.34)$$

Questi previsori devono essere non distorti, vale a dire

$$E[B_s] = b_s = \sum_{i=1}^n \lambda_{i,s} E[Z(u_i)] \quad (2.35)$$

e per la (2.31) si ha

$$b_s = \sum_{i=1}^n \lambda_{i,s} \sum_{q=1}^S b_q m_q(u_i), \quad (2.36)$$

da cui

$$b_s = \sum_{q=1}^S b_q \sum_{i=1}^n \lambda_{i,s} m_q(u_i). \quad (2.37)$$

Se la funzione  $m_s(u)$  è linearmente indipendente allora per la (2.37) si ha

$$\sum_{i=1}^n \lambda_{i,s} m_q(u_i) = \begin{cases} 1 & \text{se } q = s \\ 0 & \text{se } q \neq s \end{cases} \quad (2.38)$$

La varianza del previsore è

$$Var[B_s] = Var\left[\sum_{i=1}^n \lambda_{i,s} Z(u_i)\right] \quad (2.39)$$

solo se queste combinazioni lineari hanno una varianza finita per la quale

$$\sum_{i=1}^n \lambda_{i,s} = 0. \quad (2.40)$$

Questo permette di calcolare la varianza di previsione

$$Var[B_s] = \sum_{i=1}^n \sum_{j=1}^n \lambda_{i,s} \lambda_{j,s} \gamma(u_i - u_j), \quad (2.41)$$

che deve essere minimizzata sotto la condizione di non distorsione (2.38). Introducendo i moltiplicatori di Lagrange si arriva ad un sistema di equazioni simile a quello del *kriging* ordinario:

$$\begin{cases} \sum_{j=1}^n \lambda_{j,s} \gamma(u_i - u_j) + \mu_{0,s} + \sum_{q=1}^S \mu_{q,s} m_q(u) = 0 & \text{per } i = 1, \dots, n \\ \sum_{i=1}^n \lambda_{i,s} = 0 \\ \sum_{i=1}^n \lambda_{i,s} m_q(u_i) = \begin{cases} 1 & \text{se } q = s \\ 0 & \text{se } q \neq s \end{cases} \end{cases} \quad (2.42)$$

Risolvendo questo sistema per  $s = 1, \dots, S$  si ottengono i coefficienti  $\lambda_{i,s}$  e da questi i  $b_s$ .

## 2.9 Convalida incrociata

E' utile saggiare la bontà di un modello, per confrontarlo con altri e scegliere quello che si adatta in modo migliore al fenomeno che si studia. In geostatistica la correlazione spaziale rende difficile l'uso di test statistici.

Una procedura utilizzabile è la validazione incrociata che permette di confrontare i valori previsti con i valori veri, usando solo l'informazione che si ricava dall'insieme dei dati osservati. Per ogni sito di misurazione  $u_i$ , il valore viene temporaneamente eliminato e viene previsto come se fosse ignoto, utilizzando gli altri valori campionati. Queste previsioni si possono quindi confrontare con i valori veri, calcolando i residui tra valori osservati e valori previsti. Al fine di paragonare i modelli utilizzati e i criteri applicati, viene utilizzato un indice di convalida incrociata (*cross-validation*) dato da:

$$CV = \frac{1}{n} \sum_{i=1}^n \frac{e_i^2}{Var(e_i)},$$

dove  $e_i = Z(u_i) - \phi_i$  sono i residui tra i valori osservati  $Z(u_i)$  e i valori previsti  $\phi_i$ . Quanto più l'indice  $CV$  è vicino a 1, migliore è la previsione che si ottiene applicando una specifica combinazione di metodo e modello.



## Capitolo 3

### Analisi geostatistica

Questa tesi ha lo scopo di fornire i fondamenti metodologici per la modellazione e previsione spaziale, mediante il *software* R (Appendice D), della concentrazione di inquinanti in acque sotterranee, a partire da misure effettuate in pozzi campionati. La modellazione e previsione spaziale sono utili per indagare eventuali situazioni di degrado idrico, avendo un'accurata conoscenza del territorio e della geologia.

Temperatura ( $^{\circ}C$ )	Ferro $\mu g/l$
pH	Manganese $\mu g/l$
Conducibilità Elettrica $\mu S/cm$ a $20^{\circ}C$	Rame $\mu g/l$
Durezza Totale $mg/l$ $CaCO_3$	Zinco $\mu g/l$
Alcalinità $mg/l$ $HCO_3$	Cromo Totale $\mu g/l$
Cloruri $mg/l$	Piombo $\mu g/l$
Solfati $mg/l$ come $SO_4$	Composti Organo-alogenati Totali $\mu g/l$
Nitrati $mg/l$ come $NO_3$	Calcio $\mu g/l$
Nitriti $mg/l$ come $NO_2$	Magnesio $mg/l$
Ione ammonio $mg/l$ come $NH_4$	Potassio $mg/l$
Fosfati $\mu g/l$ come $P_2O_5$	Arsenico $\mu g/l$
Fluoruri $\mu g/l$	Cadmio $\mu g/l$
Sodio $mg/l$	Nichel $\mu g/l$

Tabella 3.2: Parametri chimici e fisici misurati

La base dati è stata fornita dall'ARPAV ( Appendice A) e contiene rilevazioni eseguite su acque sotterranee (di profondità da 4 a 380 metri) di 15 comuni

dell'alta provincia padovana<sup>1</sup> dal 12 Gennaio 1990 al 22 Agosto 1996. Potrebbe risultare poco interessante studiare concentrazioni di inquinanti relative a misurazioni effettuate in tempi così lontani, ma bisogna tenere presente che le fonti di inquinamento spesso persistono per lunghi anni e le metodologie applicate nella tesi possono essere utilizzate anche per insiemi di dati riguardanti campagne di rilevamento successive.

I dati forniti riportano le coordinate spaziali, il comune e la profondità del pozzo, la data del prelievo, i giorni dal primo prelievo e poi una serie di indicatori dell'analisi chimica e chimico-fisica relativi ai campioni di acqua raccolti (Tabella 3.2). Le sostanze chimiche rilevate non sono intrinsecamente inquinanti, ma forniscono utili indicazioni sulla qualità dell'acqua e del suolo sovrastante; la loro distribuzione nello spazio giustifica il trattarle come variabili regionalizzate, le quali assumono valori diversi a seconda della collocazione spaziale del punto in cui vengono osservate.

COMUNE	NUMERO SITI DI RILEVAZIONE
Campodoro	7
Campo San Martino	91
Carmignano di Brenta	268
Cittadella	837
Curtarolo	61
Fontaniva	226
Galliera Veneta	249
Gazzo Padovano	111
Grantorto	149
Piazzola Sul Brenta	98
San Giorgio in Bosco	254
San Martino di Lupari	389
San Pietro in Gù	232
Tombolo	434
Villafranca Padovana	37
<i>Totale</i>	<i>3443</i>

Tabella 3.4: Numero di pozzi per comune

I dati possono quindi essere visti come osservazioni spazio-temporali del

<sup>1</sup>Campodoro, Campo San Martino, Carmignano di Brenta, Cittadella, Curtarolo, Fontaniva, Galliera Veneta, Gazzo Padovano, Grantorto, Piazzola Sul Brenta, San Giorgio in Bosco, San Martino di Lupari, San Pietro in Gù, Tombolo, Villafranca Padovana

tipo  $z(t,s)$  dove i  $t_i$  indicano i giorni (da 0 a 2424) e  $s_j$  ( $j = 1, \dots, 3443$ ) i siti di rilevazione, cioè i pozzi campionati.

Per costruire le mappe di concentrazione è stato utilizzato il *software* R e sono stati seguiti questi passi:

1. **ISPEZIONE PRELIMINARE DEI DATI:** allo scopo di ricercare eventuali errori di trascrizione dei dati, dati mancanti e ridondanze. Il *software* R riconosce in modo diverso le lettere maiuscole e quelle minuscole, dunque si è dovuto uniformare la base dati. Per alcuni pozzi sono state effettuate misurazioni a diverse profondità e in più occasioni nel corso dei sette anni; in molti casi, però, mancano in parte e anche totalmente i dati relativi all'analisi chimica e chimico-fisica. Il fenomeno dei dati mancanti può essere dovuto a varie cause: omissioni involontarie, cancellazioni nella fase di elaborazione e trascrizione dei dati, impossibilità di effettuare l'analisi. In questa tesi si è deciso di escludere l'unità di cui manca il dato su una o più variabili in modo definitivo, cioè ignorandola in tutte le analisi. L'esclusione definitiva è stata applicata prima dell'analisi spaziale, poiché dà risultati più facili da interpretare, anche se comporta la rinuncia a tutte le informazioni raccolte sull'unità esclusa [2].
2. **STUDIO DELL'AREA GEOGRAFICA ESAMINATA:** è essenziale conoscere le caratteristiche del territorio e la geologia della zona interessata dalle misurazioni per effettuare la modellazione e la previsione spaziale, e per comprendere i risultati a cui si giunge. Come si può vedere nella Figura 3.1, l'area geografica interessata dalla rilevazione rientra sostanzialmente in un rettangolo di dimensioni 20 x 25 km nell'alta provincia padovana. Il fiume Brenta la attraversa da Nord-Ovest a Sud-Est. Quest'area si trova nella zona di passaggio tra l'alta pianura e la media pianura, detta fascia delle risorgive. Le risorgive si trovano lungo una ristretta fascia, al passaggio tra l'alta e la media pianura, in tutta la Pianura Padana. Esse derivano dall'affioramento della falda in aree dove la sua superficie è molto vicina al piano campagna e dove i terreni che costituiscono il sottosuolo passano da permeabili a impermeabili [11]. Seguendo il corso del Brenta, l'alta pianura va da Bassano del Grappa (150 m s.l.m.) a Fontaniva (40 m s.l.m), dove il fiume ha un letto molto ampio e sassoso, perciò le sue acque scendono verso le falde freatiche superficiali; mentre la media arriva fino a Ponte di Brenta (10 m s.l.m), dove il letto del fiume si restringe



Figura 3.1: Area geografica esaminata

e si approfondisce sul piano campagna. La linea delle risorgive, ubicata poco a sud di Cittadella, delimita la massima estensione meridionale del complesso ghiaioso dove è ospitato l'acquifero indifferenziato. Il limite settentrionale delle risorgive definisce la massima estensione della fascia delle falde in pressione. Per la collocazione cartografica dei pozzi è stato utilizzato il sistema di coordinate UTM (Universale Traversa di Mercatore): ogni punto è individuato dalle coordinate Est e Nord espresse in metri, rispetto all'origine del quadrante di appartenenza, alle quali viene anteposto un numero a seconda della zona. L'area geografica esaminata appartiene alla zona 32 nella griglia UTM. Nei dati forniti le coordinate Est, denominate Meridiano, vanno da 706540 a 725620 e le coordinate Nord, denominate Parallelo, da 5036980 a 5062980.

3. **INDIVIDUAZIONE DI DATI OMOGENEI:** si è deciso di trovare insiemi di dati omogenei all'interno della base dati, in modo che riguardassero uno stesso periodo e provenissero da pozzi che estraggono acqua dalla stessa falda. Poiché la velocità di deflusso delle acque sotterranee nell'alta pianura veneta è di alcuni metri al giorno, mentre risulta lentissima, pochi centimetri al giorno, nella media pianura, i dati riguardanti uno stesso anno possono essere considerati omogenei. Nella media pianura sono presenti più falde sovrapposte, di cui la prima libera e le sottostanti in pressione. Dall'osservazione delle stratigrafie disponibili (Appendice C), si può ragionevolmente pensare che la falda libera sia compresa in una zona che arriva fino a 30 metri di profondità, e che tra 30 e 60 metri sia presente una falda in pressione. Di conseguenza, per ogni anno di cui si dispone di misurazioni, sono stati individuati due insiemi di dati omogenei secondo la profondità. Il numero di dati riguardanti pozzi a profondità superiore a 60 metri è esiguo e non sono disponibili informazioni stratigrafiche, perciò non è stato possibile effettuare le analisi geostatistiche.
4. **DISPOSIZIONE DEI SITI:** per ogni insieme di dati omogenei è necessario avere una mappa di localizzazione dei siti rilevati e, per ciascun inquinante, bisogna identificare i pozzi in cui i valori delle concentrazioni sono più elevati o meno elevati. Inoltre in questa fase si ha una prima idea sull'eventuale presenza di relazioni lineari tra i valori e le coordinate spaziali.

5. **INDIVIDUAZIONE DI VALORI ANOMALI E ANALISI DELLE DISTRIBUZIONI:** in alcuni casi sono stati studiati i logaritmi dei dati, in modo da ridurre l'asimmetria delle distribuzioni. Infatti sebbene il variogramma non dipenda dalla distribuzione dei dati, il calcolo della semi-varianza diventa più stabile quando la distribuzione si approssima a quella Normale e la modellazione risulta meno sensibile alla presenza di dati anomali.
6. **VERIFICA DELLA CONDIZIONE DI STAZIONARIETÀ:** l'analisi di regressione permette di stimare le superfici di *trend* qualora esistano. Nei casi in cui esistevano direzioni preferenziali di sviluppo del fenomeno sono stati considerati i residui rispetto al trend e sono state applicate le tecniche del *kriging* universale.
7. **CALCOLO DEI VARIOGRAMMI E LORO MODELLAZIONE:** in primo luogo sono state ricavate le classi di distanza (*lag*) entro cui calcolare la media delle differenze quadratiche tra i valori di tutte le possibili coppie. Viene assunta come intervallo la distanza media fra tutte le possibili coppie di punti. La massima distanza tra coppie di punti è stata ridotta in modo tale da avere almeno 30 coppie di punti per ogni intervallo. I variogrammi sono stati stimati con l'utilizzo dei modelli esponenziale (2.16) e sferico (2.15). E' sempre stato utilizzato un *nugget* pari a zero, ignorando quindi l'incertezza sperimentale e assumendo perciò la bontà delle misure di concentrazione effettuate.
8. **CALCOLO DEL KRIGING:** prevalentemente si è utilizzato il *kriging* universale rispetto a quello ordinario, poiché spesso la concentrazione di inquinanti presenta una direzione preferenziale di sviluppo.

Queste sono le fasi che si sono seguite nelle applicazioni presenti nel capitolo successivo.

# Capitolo 4

## Applicazioni

### 4.1 Introduzione

Le applicazioni si riferiscono a misure di concentrazione dei principali inquinanti delle acque sotterranee quali Cloruri, Nitrati (come  $NO_3$ ) e Composti Organo-alogenati effettuate nel 1996 (dal 04/01/1996 al 22/08/1996) in pozzi di profondità minore di 30 metri e compresa tra 30 e 60 metri.

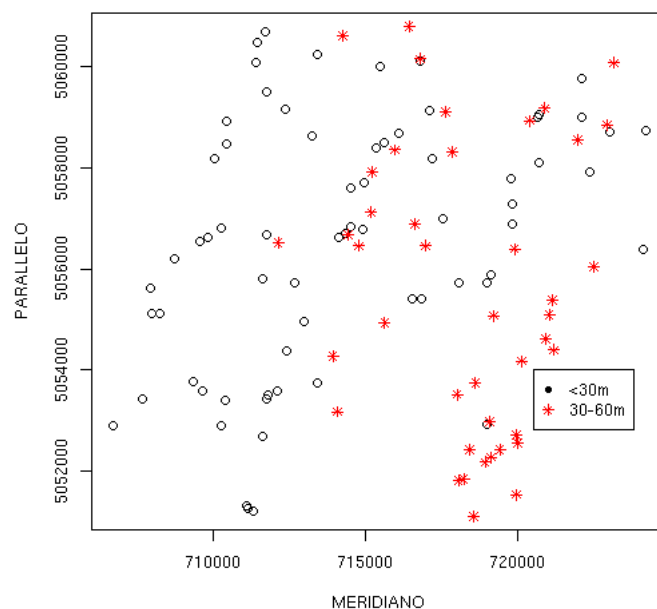


Figura 4.1: Disposizione dei 68 pozzi di profondità minore di 30 metri e dei 49 pozzi profondi tra 30 e 60 metri

L'area in cui sono stati rilevati questi dati è ristretta rispetto all'intera zona esaminata: le coordinate denominate Parallelo vanno da 5048500 a 5062120.

Dopo aver eliminato le ridondanze (dati simili rilevati nello stesso luogo e nello stesso giorno), i pozzi in cui sono state effettuate le rilevazioni sono 68 per i pozzi di profondità minore di 30 metri e 49 per quelli compresi tra 30 e 60 metri, suddivisi per comune come nelle Tabelle 4.2 e 4.4.

COMUNE	NUMERO POZZI
Carmignano di Brenta	8
Cittadella	16
Fontaniva	9
Gazzo Padovano	3
Grantorto	8
San Giorgio in Bosco	1
San Martino di Lupari	6
San Pietro in Gù	10
Tombolo	7
<i>Totale</i>	<i>68</i>

Tabella 4.2: Numero di pozzi (profondità minore 30 m) per comune

COMUNE	NUMERO POZZI
Campo San Martino	1
Carmignano di Brenta	1
Cittadella	9
Fontaniva	5
Galliera Veneta	3
Grantorto	2
San Giorgio in Bosco	18
San Martino di Lupari	7
Tombolo	3
<i>Totale</i>	<i>49</i>

Tabella 4.4: Numero di pozzi (profondità compresa tra 30 e 60 m) per comune

Per verificare una possibile relazione lineare tra le variabili studiate, sono stati ricavati i coefficienti di correlazione tra le sostanze, la profondità e le direzioni Est e Nord (Tabella 4.6 e 4.8 ).



	COMPOSTI ORGANO-ALOGENATI	CLORURI	NITRATI
CLORURI	0.30734		
NITRATI	0.37197	0.34696	
EST	0.36487	0.24224	0.69009
NORD	0.16601	-0.44079	0.27385
PROFONDITÀ	-0.00419	-0.25534	0.05890

Tabella 4.6: Correlazioni (pozzi 30 metri)

Nei pozzi a profondità minore di 30 metri, per i composti e i nitrati c'è una correlazione prossima a zero con la profondità e anche i cloruri hanno una bassa correlazione di segno negativo con la profondità. Questo supporta l'ipotesi che i dati raccolti siano sufficientemente omogenei e che provengano tutti dalla stessa falda acquifera.

C'è una correlazione lineare positiva fra elementi chimici e fra ciascuno di essi e la direzione Est, particolarmente elevata per i Nitrati, e la direzione Nord, tranne per i Cloruri con i quali è correlata negativamente. La presenza di correlazione lineare tra sostanze e direzioni indica la possibile esistenza di uno sviluppo direzionale delle sostanze.

	COMPOSTI ORGANO-ALOGENATI	CLORURI	NITRATI
CLORURI	0.8015065		
NITRATI	0.8247962	0.9580914	
EST	0.3423528	0.3568597	0.3318848
NORD	0.4798859	0.4532456	0.5189177
PROFONDITÀ	-0.6098119	-0.4130006	-0.4671998

Tabella 4.8: Correlazioni (pozzi 30-60 metri)

Nei pozzi a profondità compresa tra 30 e 60 metri si sono notate in particolare un'elevata correlazione tra le sostanze e una correlazione negativa tra le sostanze e la profondità.

Di seguito si presentano in dettaglio le fasi che portano alla previsione spaziale di cloruri, nitrati e composti organo-alogenati solamente per i pozzi di profondità minore di 30 metri, poiché non si sono rilevate particolari differenze nelle mappe di concentrazione a profondità maggiori.

## 4.2 I Cloruri

### 4.2.1 Descrizione e rappresentazione spaziale dei dati

E' stato scelto di studiare i Cloruri poiché sono delle sostanze riscontrabili naturalmente nelle acque, ma aumentano in presenza di scarichi civili e industriali. Effetti avversi alla salute si possono verificare a concentrazioni uguali o superiori a 200 mg/l. Nella prima fase di lavoro si valuta se i dati sono disposti in modo da rivelare la loro struttura spaziale.

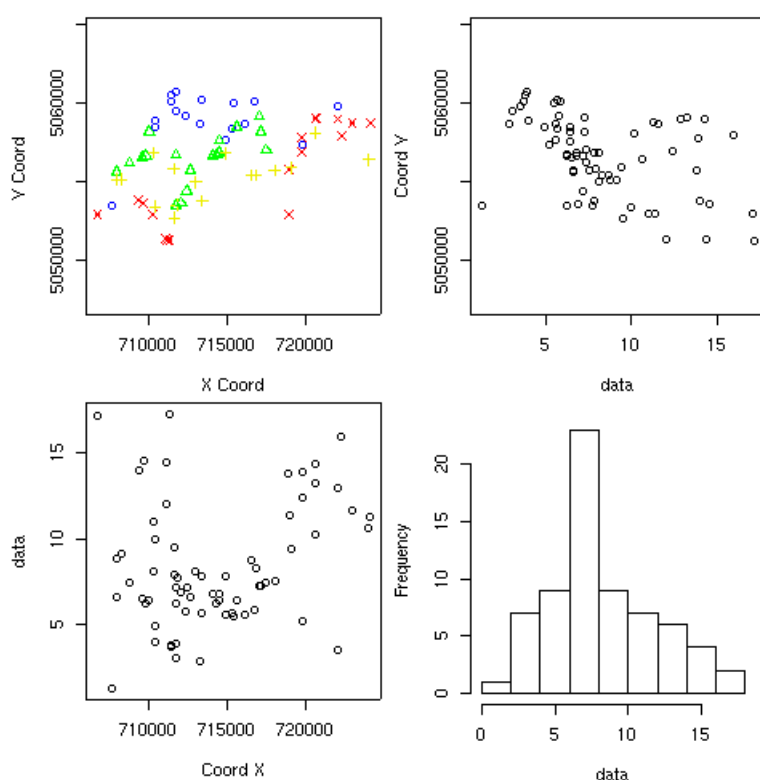


Figura 4.2: Disposizione dei siti e istogramma dei dati

Grazie alla figura 4.2, in cui viene presentata la mappa di localizzazione dei siti di prelievo, si può analizzare la distribuzione della concentrazione di cloruri nel territorio, identificando i punti in cui ci sono i valori più elevati. Il *software* ha scelto automaticamente di dividere i dati in quattro gruppi, rappresentando i valori più bassi in blu e quelli più elevati in rosso.

La mappa di localizzazione ci permette di vedere che l'aumento dei valori di cloruro sembra seguire la direzione in cui scorre il fiume Brenta, e allo stesso tempo aumenta nei siti più lontani da esso, sia in direzione Nord-Est che Sud-

Ovest, pur rimanendo ampiamente sotto alla soglia di pericolosità per la salute, come si può notare anche dalla figura (4.3). Nella figura 4.2 il secondo e terzo riquadro rappresentano i dati rispetto alle coordinate dei pozzi e permettono di individuare l'eventuale presenza di relazione lineare tra le coordinate spaziali e i valori di concentrazione. I valori più bassi di cloruro sono facilmente individuabili a nord, all'altezza della linea delle risorgive, e aumentano nella zona dell'acquifero in pressione, si può ipotizzare la presenza di un *trend* in direzione Nord-Sud.

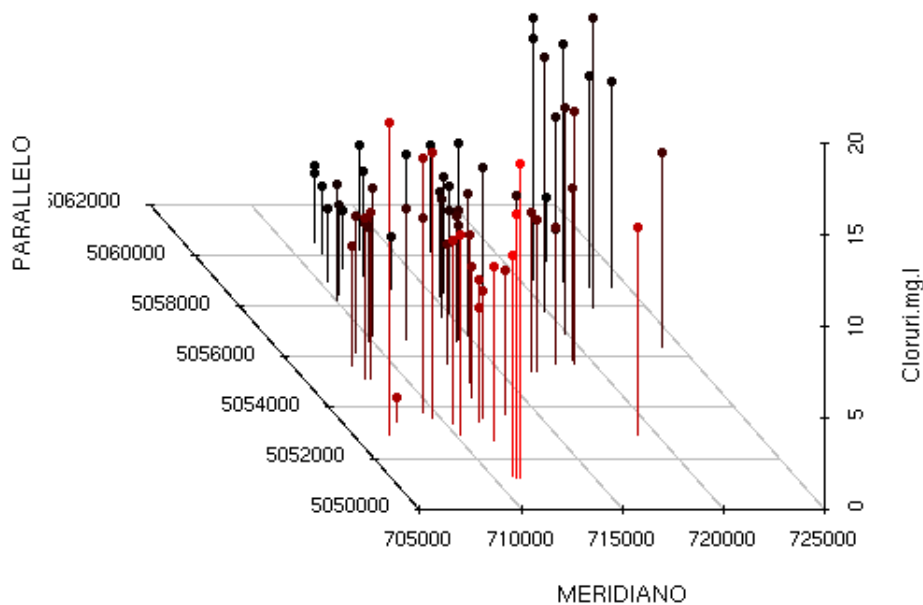


Figura 4.3: *Scatterplot* della concentrazione di cloruro

Sempre in figura 4.2 si può osservare l'istogramma in cui non viene considerata la dimensione spaziale del fenomeno: si nota che la distribuzione dei dati non si discosta eccessivamente dalla Normale, c'è una classe modale al centro della distribuzione e non si identificano valori anomali.

Per supportare quanto detto, di seguito si presentano delle statistiche di sintesi: alcune misure di posizione (media, mediana, minimo, massimo e quantili), la deviazione standard come misura di variabilità, l'indice di curtosi, che misura la densità dei dati attorno alla propria media (vale 3 per la Normale,

distribuzione di riferimento) e l'indice di asimmetria di Pearson (è uguale a 0 per distribuzioni simmetriche) (Tabella 4.10).

CLORURI	
MINIMO	1.300 mg/l
PRIMO QUARTILE	6.125 mg/l
MEDIANA	7.400 mg/l
MEDIA	8.357 mg/l
TERZO QUARTILE	10.700 mg/l
MASSIMO	17.200 mg/l
DEVIAZIONE STANDARD	3.607 mg/l
INDICE DI CURTOSI	2.77
INDICE DI ASIMMETRIA	0.63
Numero osservazioni	68

Tabella 4.10: Statistiche di sintesi: misure di posizione, di variabilità e di forma

#### 4.2.2 Stima del variogramma

Per analizzare la dipendenza spaziale si procede con la costruzione e stima del semivariogramma isotropo (2.20) che rappresenta appunto la relazione tra variabilità spaziale (*semivarianza*) e distanze vettoriali  $h$ .

Classe	$N(h)$	$h$	$\hat{\gamma}(h)$
1	211	2498.455	7.555664
2	285	3747.683	6.608789
3	347	4996.910	8.722363
4	301	6246.138	11.664917
5	293	7495.365	15.000939
6	198	8744.593	18.793889
7	192	9993.820	22.380156
8	127	11243.048	20.526575
9	91	12492.275	17.880879
10	65	13741.503	15.010846
11	50	14990.730	20.394600

Tabella 4.12: Stima del semivariogramma onnidirezionale

Dalla disposizione dei siti rilevati si ha come massima distanza tra i pozzi 18448.73 metri, ma si è deciso di ridurla a 15000 metri per avere, in ogni classe di distanze  $h$  (*lag*), un numero di coppie di siti  $N(h)$  maggiore di 30, numero ragionevole per ottenere una buona stima del variogramma.

La minima distanza tra i pozzi è 2498.455 metri, al di sotto della quale non si può stimare il semivariogramma.

Inizialmente viene rappresentato il variogramma empirico attraverso la costruzione della nuvola del variogramma. All'interno della nuvola i punti sono disegnati per ogni coppia di siti  $u_i$  e  $u_j$ . La distanza tra le coppie è sull'asse delle ascisse, mentre le ordinate rappresentano la stima della varianza.

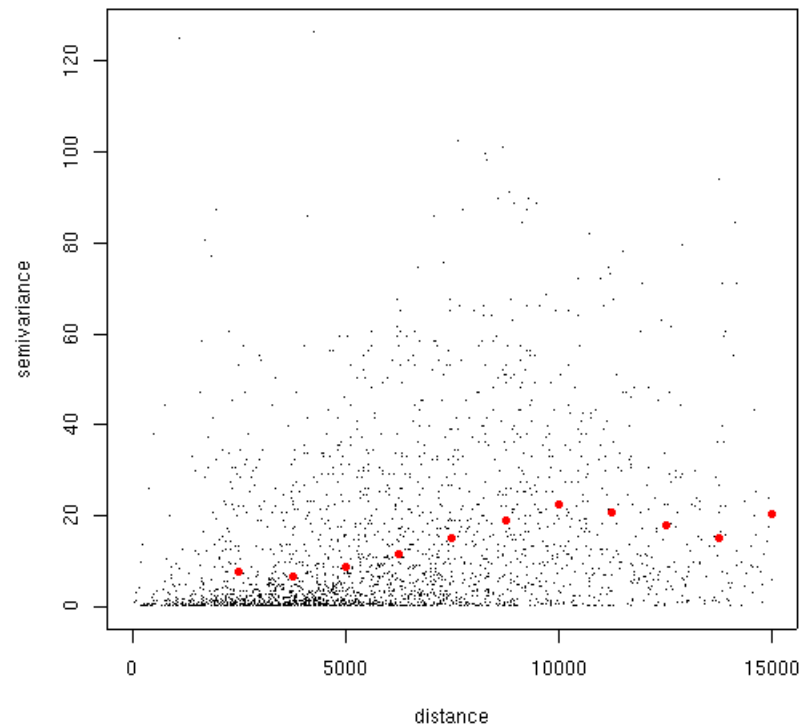


Figura 4.4: Nuvola del variogramma e stima del variogramma isotropo

Nella figura 4.4 si riporta la nuvola del variogramma e la stima del variogramma isotropo stimato sugli 11 valori del lag.

Si può notare un comportamento abbastanza regolare e continuo.

L'utilizzo del variogramma robusto di Cressie e Hawkins (2.21) può essere vantaggioso per ridurre l'effetto di valori anomali senza rimuovere osservazio-

ni dall'insieme dei dati. In questo caso non si notano particolari differenze tra i due semivariogrammi, dunque in seguito si farà riferimento allo stimatore classico.

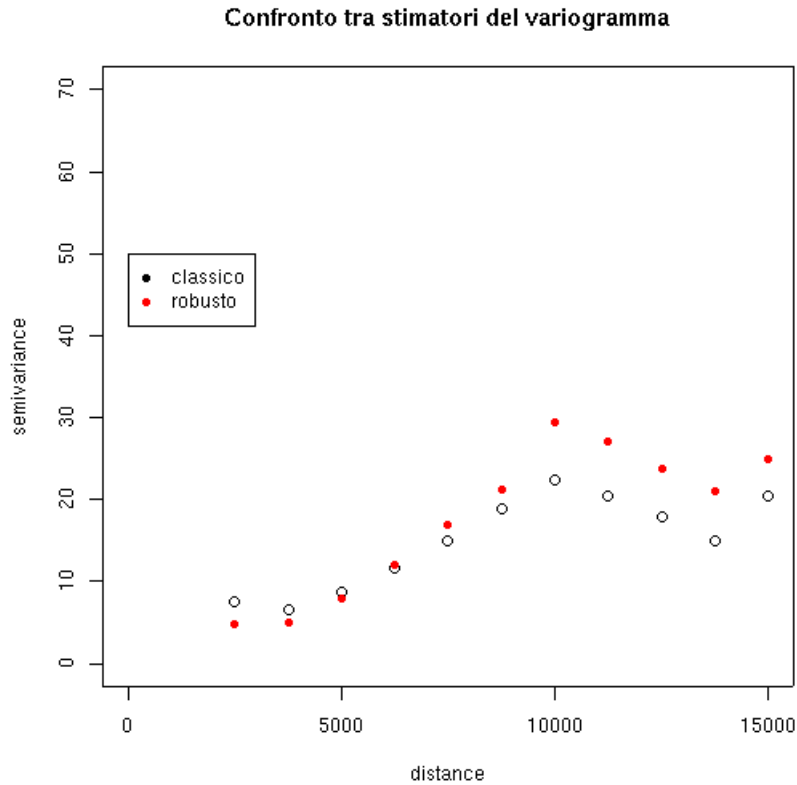


Figura 4.5: Confronto tra stimatore classico e robusto

Dato che si ha un numero sufficientemente elevato di coppie di punti per ogni valore del *lag*, sono stati stimati i semivariogrammi direzionali per approfondire l'indagine.

Nella figura 4.6 si riportano 4 semivariogrammi direzionali ( $\alpha = 0^\circ \pm 22.5^\circ$ ,  $\alpha = 45^\circ \pm 22.5^\circ$ ,  $\alpha = 90^\circ \pm 22.5^\circ$ ,  $\alpha = 135^\circ \pm 22.5^\circ$ ). Sono utili per mettere in evidenza il *trend* e per individuare l'anisotropia, che è rappresentata da un variogramma con un *range* più elevato nella direzione di massima continuità (anisotropia geometrica) oppure in differenze nel valore del *sill* (anisotropia zonale). In quest'ultimo grafico si può notare che per distanze inferiori ai 10 km i variogrammi hanno lo stesso comportamento, mentre per distanze maggiori il semivariogramma nella direzione individuata dall'angolo polare  $\alpha = 0^\circ$  aumenta e per  $\alpha = 45^\circ$  ha maggiore variabilità.

Non si è riscontrata anisotropia geometrica, per cui non si è ritenuto op-

portuno ricondurre il variogramma anisotropo ad uno isotropo tramite una trasformazione.

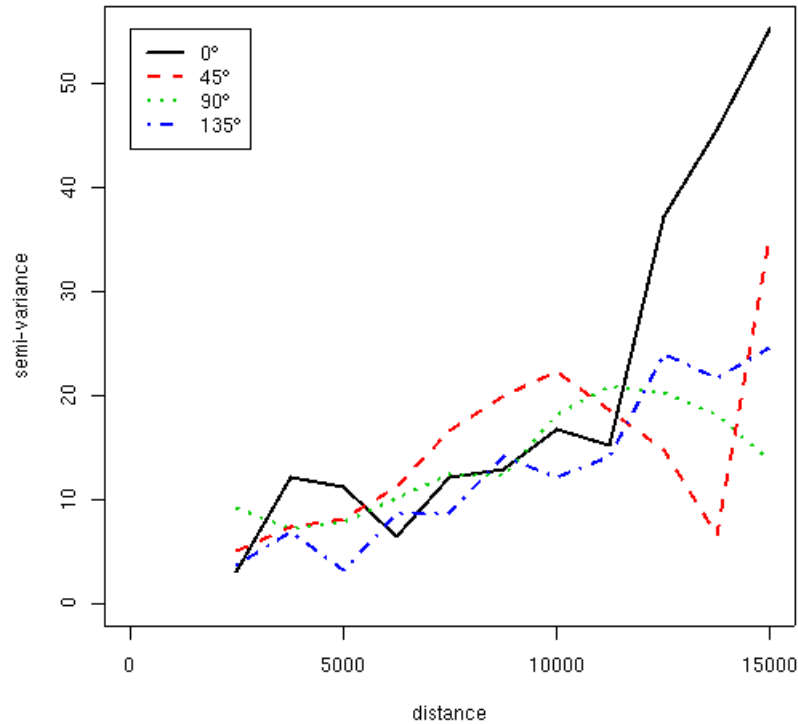


Figura 4.6: Semivariogrammi direzionali

### 4.2.3 Costruzione del modello - Stima dei parametri

Successivamente si è cercato un modello per descrivere la correlazione spaziale conoscendo il variogramma empirico. Nel nostro caso, avendo riscontrato un comportamento anisotropo del variogramma, l'ipotesi  $E(Z(u)) = m, \forall u$ , non è verificata, dunque non verrà utilizzato il modello di funzione aleatoria stazionaria. Consideriamo il metodo che presuppone l'esistenza di un *trend*: il *kriging* universale.

La direzione preferenziale di sviluppo individuata non siamo certi che corrisponda a caratteristiche peculiari del fenomeno, ma sappiamo che i cloruri presenti naturalmente nella pianura veneta sono imputabili alla vicinanza del mare e possiamo supporre che diminuiscano in ragione della maggiore lontananza da esso.

Considereremo il variogramma dei residui  $Y(u) = Z(u) - \widehat{m}(u)$ , in cui il

trend  $\hat{m}(u)$  può essere un polinomio di primo grado delle coordinate  $x_1$  e  $x_2$  oppure un polinomio di secondo grado. Nel nostro caso le coordinate  $x_1$  e  $x_2$  corrispondono a Meridiano e Parallelo. La metodologia utilizzata per calcolare il trend è dunque la regressione lineare e i coefficienti sono calcolati con il metodo dei minimi quadrati. Si ottengono i seguenti risultati.

Per il polinomio di primo grado:

	Estimate	Std. Error	t value	Pr(>  t )
(Intercept)	4429	746.7	5.932	1.27e-07
MERIDIANO	4.005e-04	8.400e-05	4.768	1.09e-05
PARALLELO	-9.309e-04	1.520e-04	-6.123	5.93e-08

con coefficiente di determinazione  $R^2 = 0.403$ , che rappresenta la frazione di devianza spiegata dal modello.

Per il polinomio di secondo grado:

	Estimate	Std. Error	t value	Pr(>  t )
(Intercept)	1.843e+06	1.323e+06	1.393	0.1686
MERIDIANO	-3.690e-01	2.217e-01	-1.664	0.1011
PARALLELO	-6.760e-01	5.309e-01	-1.273	0.2076
I(MERIDIANO^2)	3.559e-08	1.892e-08	1.881	0.0646
I(PARALLELO^2)	6.232e-08	5.349e-08	1.165	0.2485
MERIDIANOxPARALLELO	6.298e-08	4.650e-08	1.354	0.1805

con coefficiente di determinazione  $R^2 = 0.5132$ .

Da questa analisi il secondo modello risulta maggiormente significativo, infatti ha un coefficiente di determinazione più alto (sebbene  $R^2$  non sia mai molto elevato con dati di tipo ambientale), ma l'apporto delle variabili alla spiegazione del modello è meno significativo rispetto al primo, infatti hanno un *p-value* (livello di significatività osservato) più elevato e soprattutto l'interazione tra le coordinate risulta poco significativa. Per questo motivo si è deciso di utilizzare come *trend* il polinomio di primo grado.

Dunque  $\hat{m}(u) = 4429 + 0.0004005x_1 - 0.0009309x_2$ .

A questo punto vengono considerati i residui stimati  $\hat{y}(u) = Z(u) - \hat{m}(u)$  e si cerca di trovare indicazioni sui valori di *sill*, *range* e *nugget* per individuare il modello teorico che interpreti in modo migliore la variabilità spaziale. Le funzioni che si sono utilizzate per modellare il variogramma sono state il modello



esponenziale e il modello sferico. In figura 4.7 è rappresentato il semivariogramma sferico stimato con i minimi quadrati generalizzati (MQG) e con la massima verosimiglianza (MV).

	ESPONENZIALE		SFERICO	
	M V	M Q G	M V	M Q G
<i>sill</i>	7.471	11.3135	7.362	10.1026
<i>range</i>	349.9	4751.833	681.9	8518.7926
<i>nugget</i>	0	0	0	0

Il modello teorico che si adatta in modo migliore al variogramma dei residui è il modello sferico.

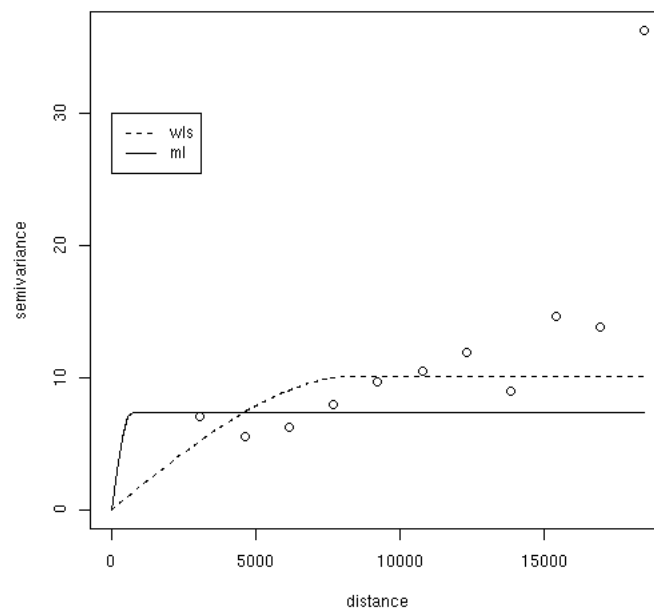


Figura 4.7: Variogramma dei residui

#### 4.2.4 Previsione e rappresentazione grafica della concentrazione di Cloruri

Per prevedere i valori di cloruri nella zona di interesse viene applicato il *kriging* universale, che tiene conto del *trend* rilevato. Il modello di semivariogramma utilizzato è quello stimato con i minimi quadrati pesati.

In ogni punto di una griglia regolare sono stati previsti i valori di cloruri, risolvendo il sistema del *kriging* universale (figura 4.8).

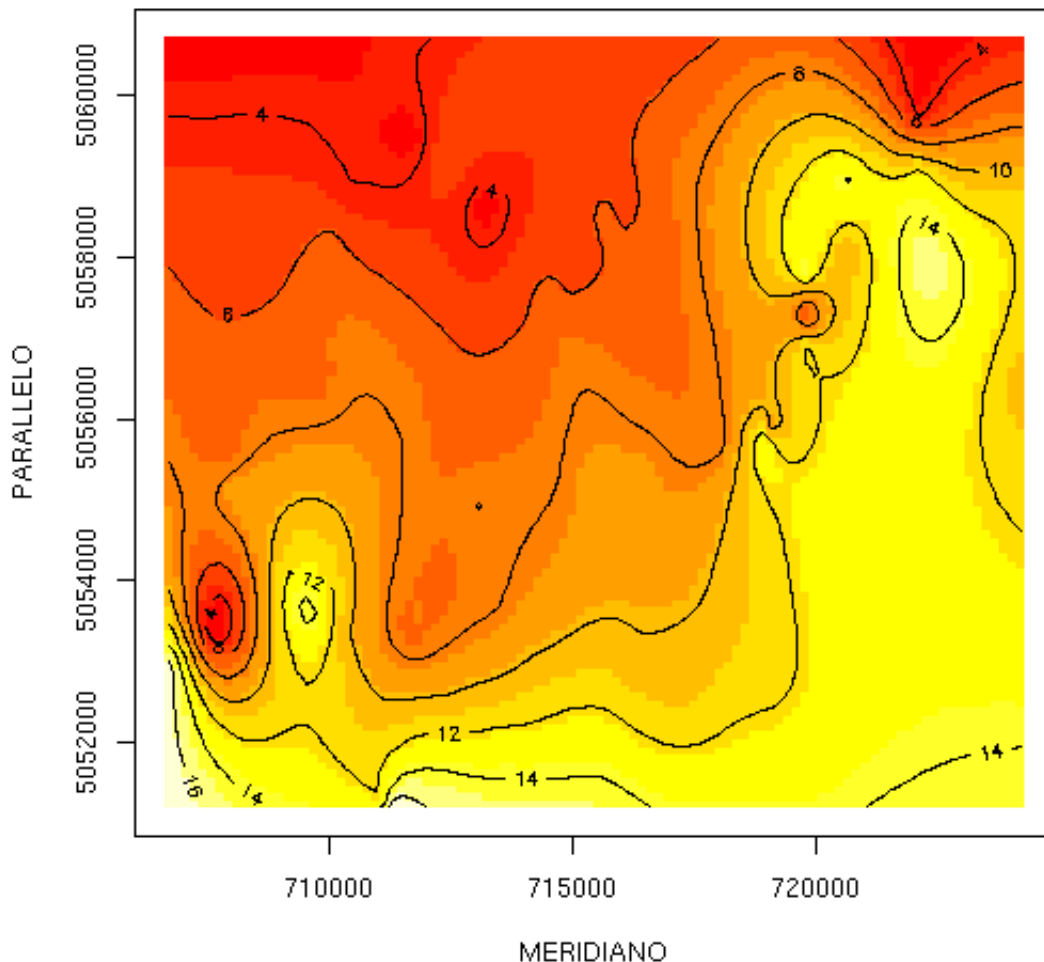


Figura 4.8: *Kriging* universale

Si può notare come la concentrazione di cloruri aumenti scendendo in direzione sud-est; nell'area studiata si possono distinguere nettamente due fasce: una che parte da nord-ovest con una concentrazione di 4 mg/l e arriva fino a 8 mg/l; l'altra dai 12 mg/l arriva a sud-est fino a 14 mg/l. Tali concentrazioni non sono sufficienti per parlare di inquinamento, si può pensare che effettivamente siano di origine naturale e derivino dalla vicinanza del mare, come avevamo supposto in precedenza, e che diminuiscano con l'aumentare della distanza da esso.

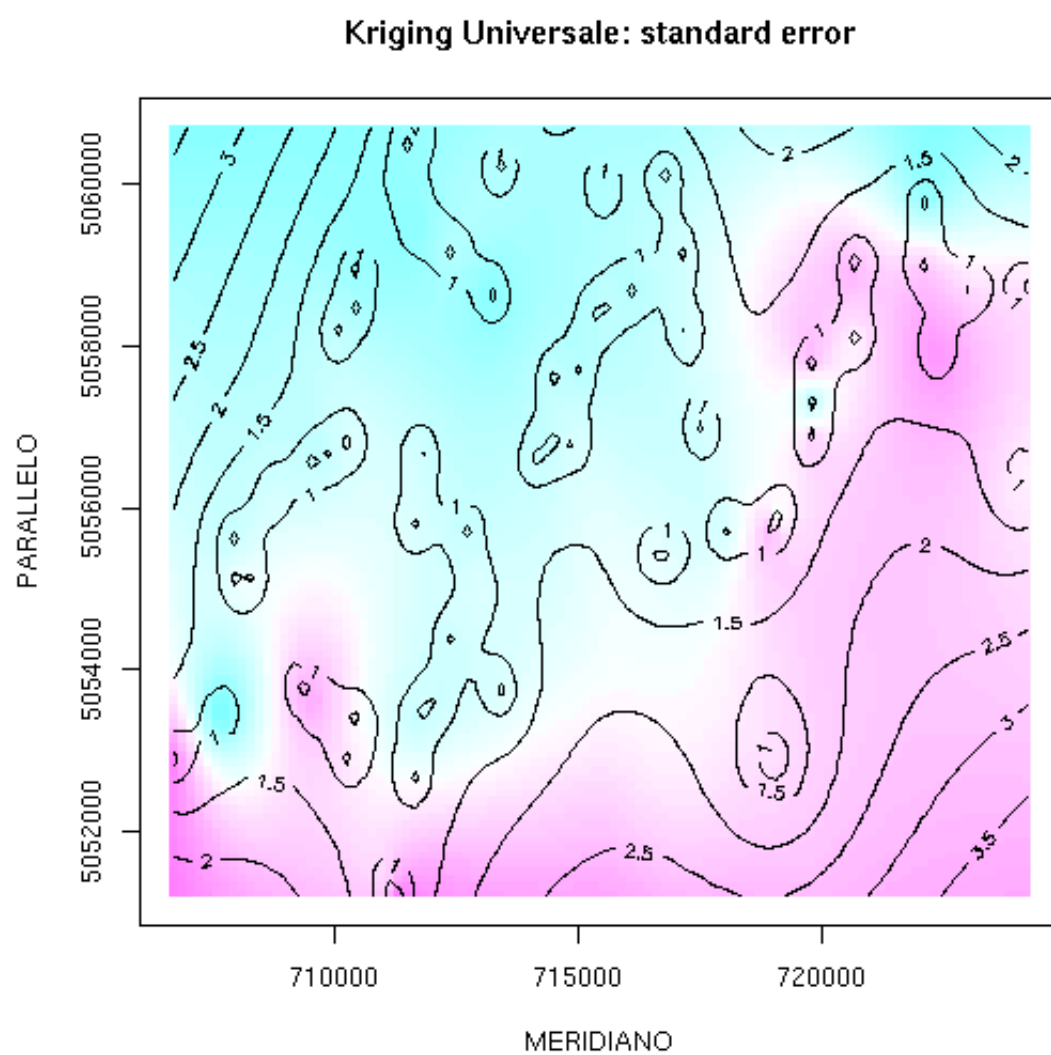


Figura 4.9: *Standard error* relativo ai valori previsti di cloruro

Di seguito vengono presentate anche le curve di livello dello scarto quadratico medio, che danno un'idea dell'andamento dell'errore di previsione (figura 4.9). Infatti una caratteristica peculiare del kriging è quella di fornire, oltre alla previsione, anche la varianza di previsione. Le mappe dell'errore di previsione possono essere utili per ottimizzare il campione: dove la deviazione standard ha valori più elevati sarà possibile aggiungere pozzi da cui trarre informazioni e se ne potranno eliminare nelle zone in cui è più bassa.

### 4.2.5 Convalida del modello

L'analisi geostatistica si conclude con la convalida incrociata mediante la quale si valuta la bontà del modello utilizzato per il *kriging*. Dalle 68 osservazioni ne viene eliminata una alla volta, al posto della quale viene previsto il valore di cloruro a partire dalle 67 rimanenti ed utilizzando lo stesso modello di variogramma. In questo modo si possono confrontare i valori previsti con i valori veri.

Per paragonare i modelli e i criteri applicati, viene utilizzato l'indice di convalida incrociata  $CV$ , che in questo caso risulta essere uguale a 0.9853, dunque si è in presenza di una leggera sottostima.

## 4.3 I Nitrati

### 4.3.1 Descrizione dei dati

I nitrati sono ioni che fanno parte del ciclo dell'azoto che si svolge in atmosfera e nel terreno, a causa della loro elevata solubilità migrano facilmente nelle falde. Nelle acque superficiali e sotterranee i livelli naturali di nitrati sono di pochi milligrammi per litro. Un aumento della concentrazione dei nitrati nell'acqua è spesso associato all'attività agricola (uso di liquami zootecnici come fertilizzanti). Un valore guida (5 mg/l) per i nitrati è stato stabilito per prevenire la metaemoglobinemia infantile, nella quale viene compromessa la capacità del sangue dei lattanti di trasportare ossigeno (in casi estremi con esito fatale). L'inquinamento da nitrati deriva da sorgenti di tipo diffuso ed è legato all'attività agricola.

Nelle figure 4.10 e 4.11 si notano valori particolarmente elevati di nitrati e la presenza evidente di un *trend*: la sostanza indagata si trova in concentrazioni

maggiori nella zona più a Est dell'area interessata dalla rilevazione, come si era dedotto in precedenza dal coefficiente di correlazione molto elevato tra la sostanza e la direzione Est ( $\rho = 0.69009$ ).

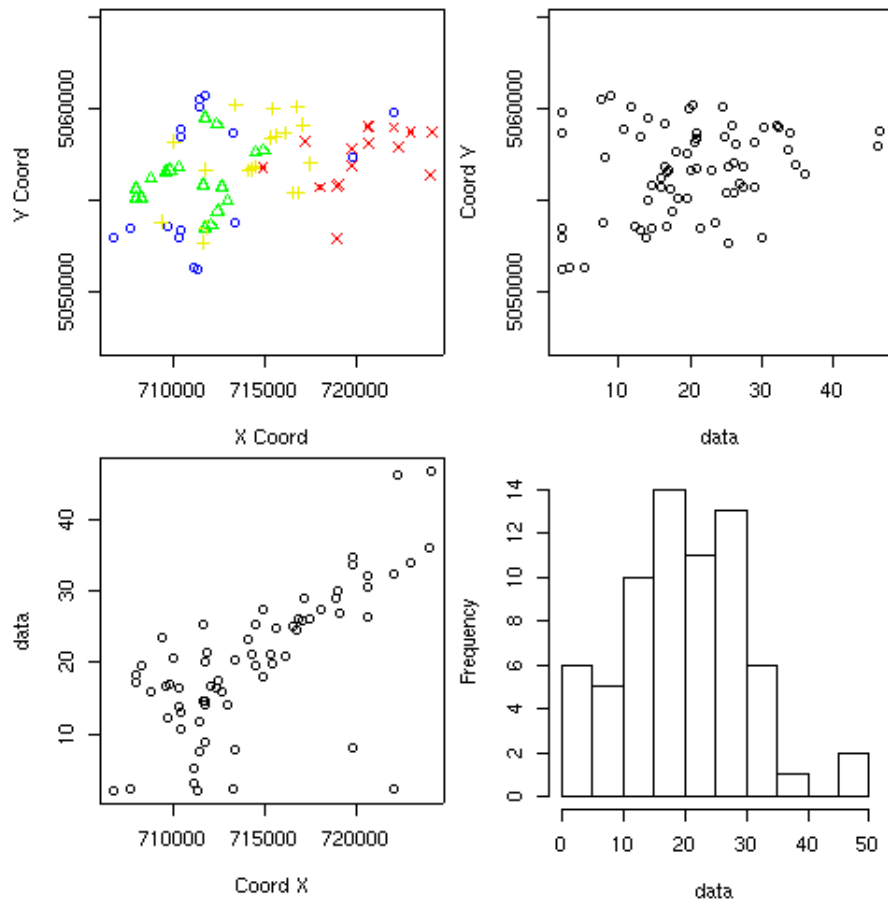


Figura 4.10: Disposizione dei siti e istogramma

L'istogramma, in cui non viene considerata la dimensione spaziale del fenomeno, presenta la distribuzione tipica di una mistura di Normali dovuta proprio alla presenza di *trend*. Il *trend* è in direzione ovest-est: sulla destra Brenta ci sono le concentrazioni più basse (2.10 mg/l), mentre sulla sinistra ci sono quelle più elevate, che arrivano fino a 46.60 mg/l.

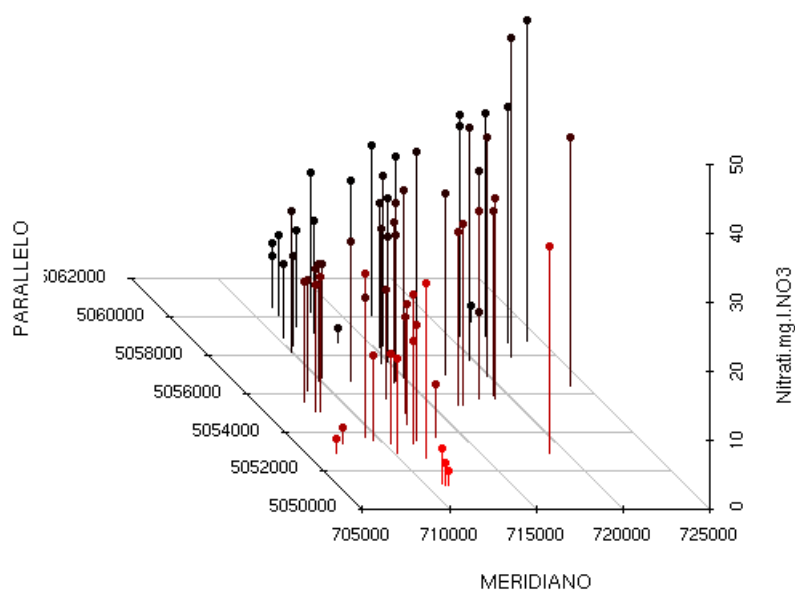


Figura 4.11: *Scatterplot* della concentrazione di nitrati

NITRATI ( $NO_3$ )	
MINIMO	2.10 mg/l
PRIMO QUARTILE	13.95 mg/l
MEDIANA	19.65 mg/l
MEDIA	19.78 mg/l
TERZO QUARTILE	26.20 mg/l
MASSIMO	46.60 mg/l
DEVIAZIONE STANDARD	9.942 mg/l
CURTOSI	3.12
ASIMMETRIA	0.25
Numero osservazioni 68	

Tabella 4.17: Statistiche di sintesi: misure di posizione, di variabilità e di forma

Dunque la variabilità del fenomeno è particolarmente elevata, come risulta dalla tabella 4.17.

### 4.3.2 Stima del variogramma

Si procede con la stima del semivariogramma isotropo. Le distanze vettoriali  $h$  e il numero di coppie di siti  $N(h)$  sono le stesse calcolate per la stima del

variogramma della concentrazione di Cloruri. Dato che si ha un numero sufficientemente elevato di coppie di punti per ogni valore del *lag* e avendo riscontrato una direzione preferenziale di sviluppo del fenomeno, utilizziamo tali informazioni per stimare i semivariogrammi anisotropi.

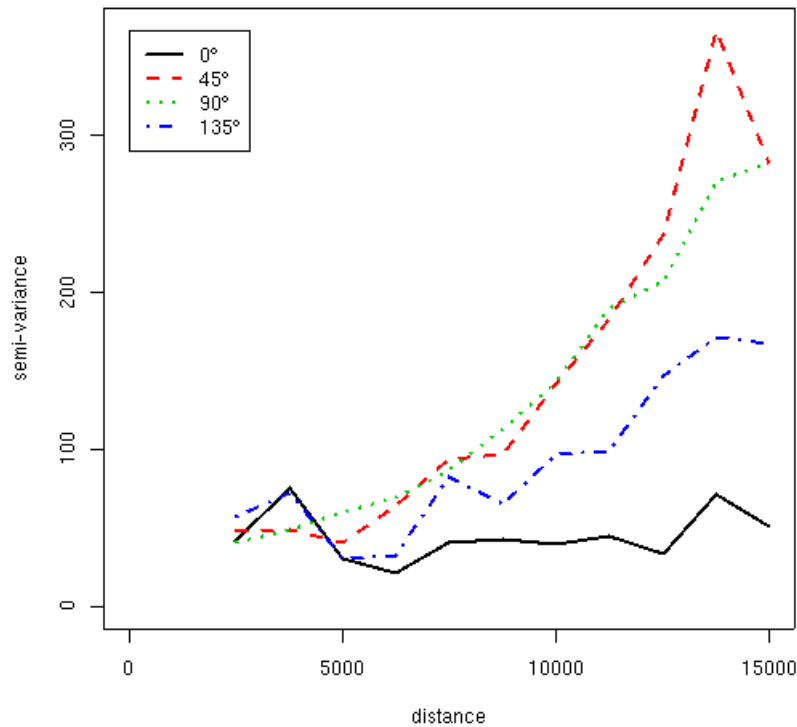


Figura 4.12: Semivariogrammi direzionali

Nella figura 4.12 si riportano i semivariogrammi direzionali ( $\alpha = 0^\circ \pm 22.5^\circ$ ,  $\alpha = 45^\circ \pm 22.5^\circ$ ,  $\alpha = 90^\circ \pm 22.5^\circ$ ,  $\alpha = 135^\circ \pm 22.5^\circ$ ). In quest'ultimo grafico si può notare la presenza di *trend*, per il comportamento parabolico dei semivariogrammi nelle direzioni sud-ovest nord-est ( $\alpha = 45^\circ$ ) e ovest-est ( $\alpha = 90^\circ$ ), come già rilevato in precedenza. Nella direzione in cui si ha minore variabilità, in questo caso  $\alpha = 0^\circ$ , si rileva il maggiore grado di correlazione dei dati, perciò verso di essa si trova il maggiore valore di *range*, ossia la maggiore distanza oltre la quale i dati si possono ritenere incorrelati. In genere, come in questo caso, le direzioni più variabili e quelle meno variabili sono tra loro ortogonali. La significativa anisotropia rilevata può essere giustificata dalla conformazione del terreno in cui si trova la falda, ma potrebbe anche essere dovuta ad una raccolta dei dati avvenuta in modo preferenziale non casuale.

### 4.3.3 Costruzione del modello

Anche in questo caso è presente un trend e quindi l'ipotesi  $E(Z(u)) = m, \forall u$ , non è verificata ed è necessario utilizzare il modello di funzione aleatoria non stazionaria.

E' stato considerato il variogramma dei residui  $Y(u) = Z(u) - \hat{m}(u)$ , e il trend  $\hat{m}(u)$ , polinomio delle coordinate  $x_1$ = Meridiano e  $x_2$ = Parallelo, è stato stimato con il metodo dei minimi quadrati.

Per il polinomio di primo grado:

	Estimate	Std. Error	t value	Pr(>  t )
(Intercept)	-1016	1.928e+03	-0.527	0.600
MERIDIANO	0.00153	2.169e-04	7.057	1.38e-09
PARALLELO	-1.137e-05	3.925e-04	-0.029	0.977

con coefficiente di determinazione  $R^2 = 0.4762$ .

Per il polinomio di secondo grado:

	Estimate	Std. Error	t value	Pr(>  t )
(Intercept)	-1.243e+07	3.337e+06	-3.725	0.000424
MERIDIANO	3.838e-01	5.592e-01	0.686	0.495088
PARALLELO	4862	1339	3.631	0.000574
I(MERIDIANO^2)	5.570e-08	4.771e-08	1.167	0.247514
I(PARALLELO^2)	-4.743e-07	1.349e-07	-3.516	0.000826
MERIDIANOxPARALLELO	-9.137e-08	1.173e-07	-0.779	0.438910

con coefficiente di determinazione  $R^2 = 0.5924$ .

Come per i cloruri, il secondo modello ha un coefficiente di determinazione più elevato, ma si è preferito utilizzare il polinomio di primo grado, poiché l'apporto dell'interazione tra le coordinate non è significativo.

Dunque utilizziamo il polinomio di primo grado:  $\hat{m}(u) = -1016 + 0.00153x_1 - 0.00001137x_2$ .

Per i Nitrati è stato studiato anche un modello che tenesse conto della posizione, rispetto al fiume Brenta, dei pozzi rilevati. L'introduzione di questa variabile *dummy* non migliora il modello e il coefficiente non risulta significativo, perciò non verrà in seguito considerato.

A questo punto vengono considerati i residui stimati  $\hat{y}(u) = Z(u) - \hat{m}(u)$  e si cercano indicazioni su *sill*, *range* e *nugget* per individuare il modello teorico.



Il modello teorico che meglio si adatta al variogramma dei residui è il modello esponenziale. In figura 4.13 è rappresentato il semivariogramma stimato con i minimi quadrati generalizzati (MQG).

	MODELLO ESPONENZIALE	
	M V	M Q G
<i>sill</i>	50.47	75.9062
<i>range</i>	331	4433.5302
<i>nugget</i>	0	0

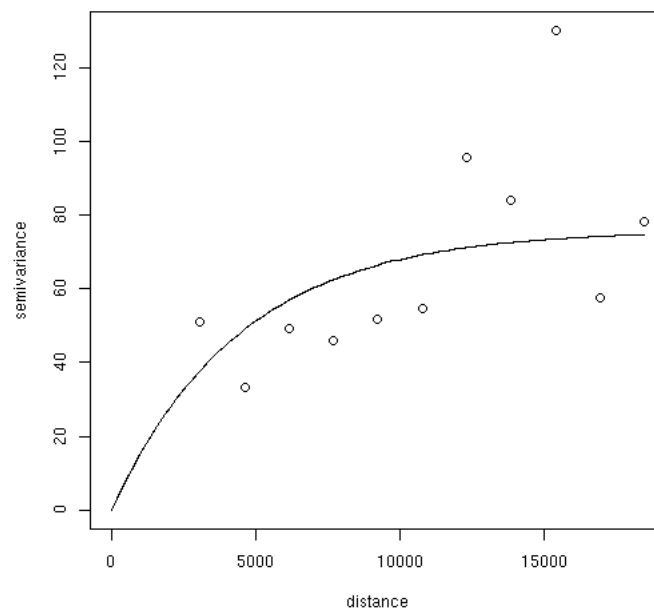


Figura 4.13: Modello esponenziale variogramma residui

#### 4.3.4 Previsione e rappresentazione grafica della concentrazione di Nitrati

Per prevedere i valori di Nitrati nella zona di interesse viene applicato il *kriging* universale, che tiene conto del *trend* rilevato.

Il modello di semivariogramma utilizzato è quello stimato con i minimi quadrati pesati, che si adatta in modo migliore al variogramma dei residui. In ogni punto della griglia sono stati stimati i valori di Nitrati, risolvendo il sistema del *kriging* universale (figura 4.14).

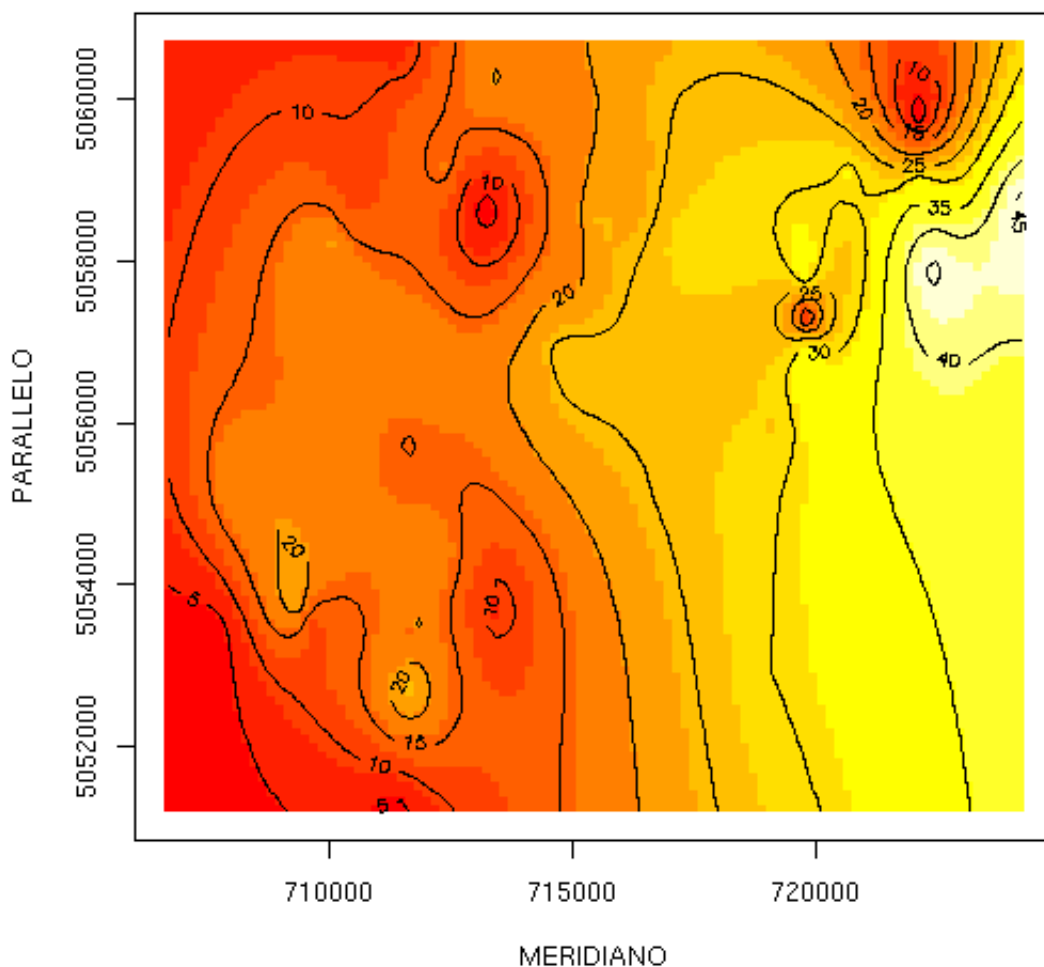


Figura 4.14: *Kriging* universale per i nitrati

Di seguito vengono presentate anche le curve di livello dello scarto quadratico medio, che danno un'idea dell'andamento dell'errore di stima (figura 4.15), e il *kriging* universale in una rappresentazione tridimensionale (figura 4.16) per una visione più realistica del fenomeno. E' evidente una divisione della zona in due aree: una a ovest con concentrazioni intorno ai 10 mg/l, con alcuni picchi di 20 mg/l, e una a est dove i valori previsti arrivano a 45 mg/l. Si può notare che i prelievi dei dati sono stati effettuati dal 04/01/1996 al 22/08/1996, dunque in un periodo dell'anno in cui vengono ampiamente utilizzati i fertilizzanti per le coltivazioni. Questa è probabilmente la causa di concentrazioni così elevate di nitrati in pozzi a profondità minore di 30 metri.

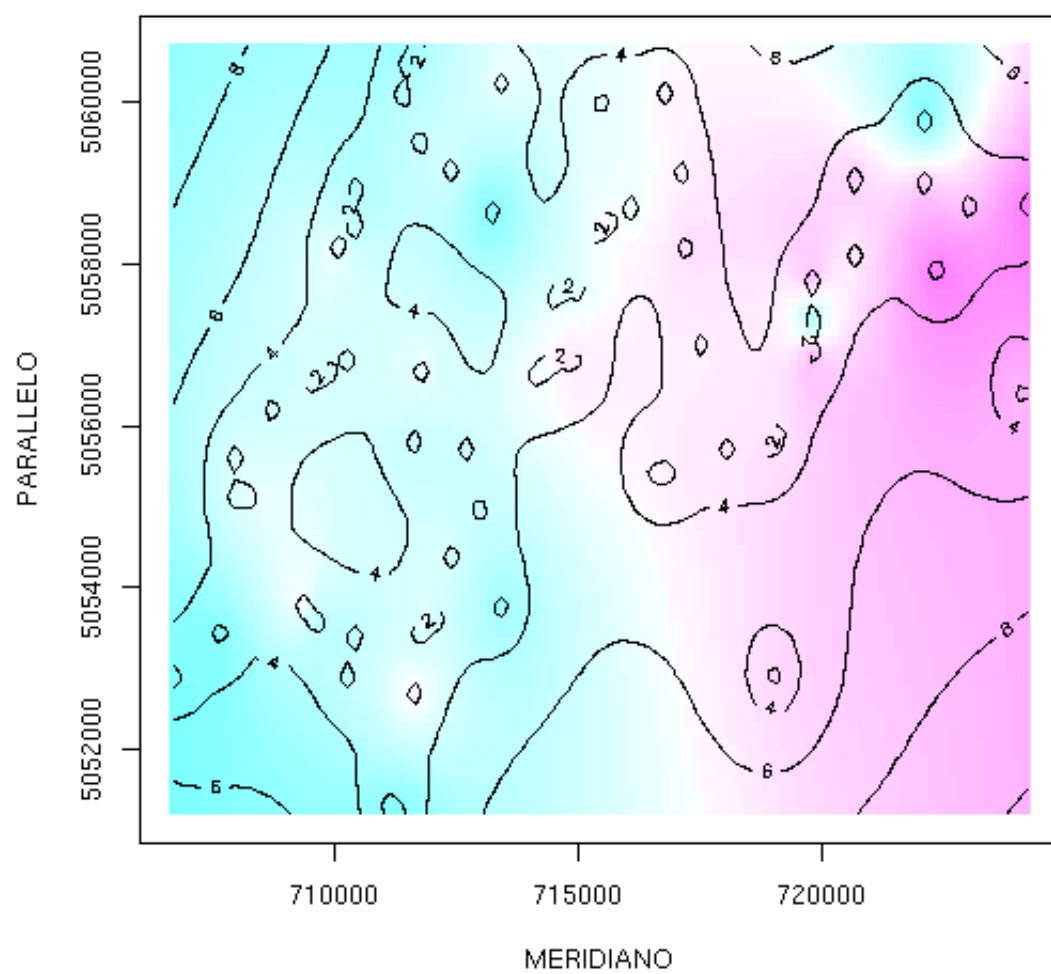


Figura 4.15: *Standard error* relativo ai valori stimati dei nitrati

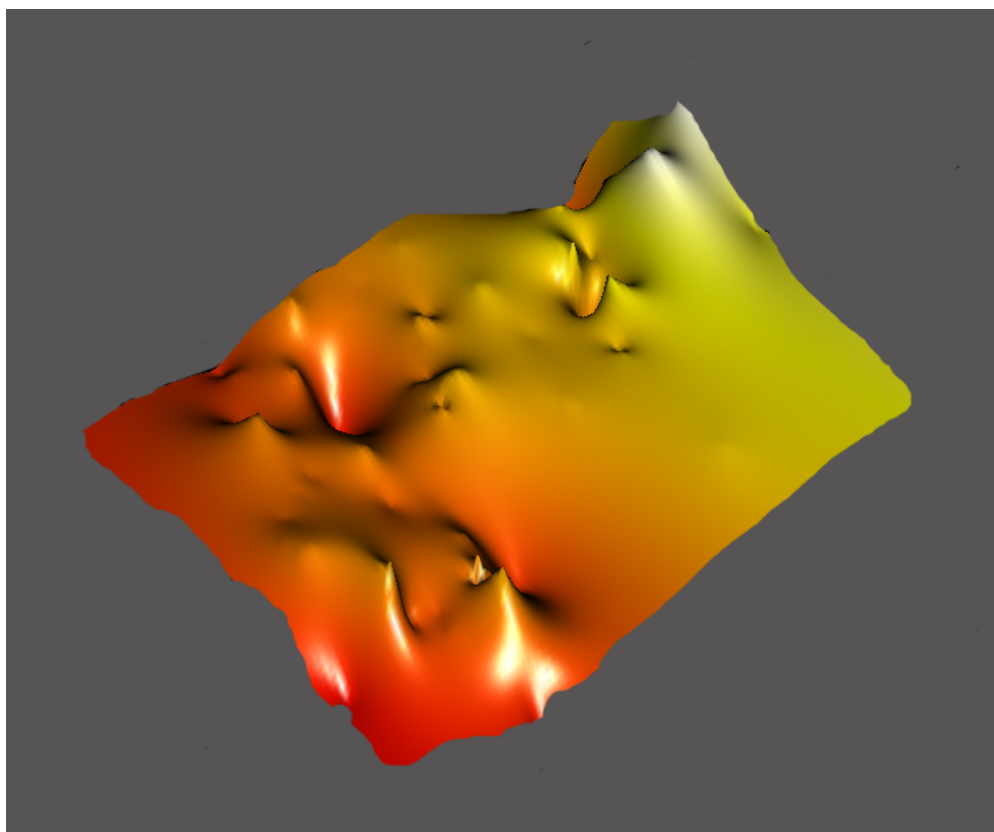


Figura 4.16: *Kriging* universale (vista tridimensionale direzione sud-ovest)

### 4.3.5 Convalida del modello

La convalida incrociata permette la valutazione della bontà del modello utilizzato per il *kriging*. Per paragonare i modelli e i criteri applicati, viene utilizzato l'indice di convalida incrociata  $CV$ , che in questo caso risulta essere uguale a 0.9855, dunque si è in presenza di una leggera sottostima.

## 4.4 I Composti Organo-alogenati

### 4.4.1 Descrizione dei dati

I solventi organo-alogenati, che abbiamo visto essere i più diffusi inquinanti delle acque sotterranee, sono una serie di composti organici (cloroformio, tricloroetilene, tetracloroetilene, 1,1,1 tricloroetano) contenenti all'interno della propria molecola un alogeno (Cl, F, Br, I). Vengono utilizzati come solventi o come materie prime in svariati processi industriali e artigianali (sgrassaggio dei

metalli, lavanderie, conerie, usi domestici, adesivi...). Questi composti hanno caratteristiche che li rendono potenzialmente molto inquinanti: infatti, nel sottosuolo, sono pressoché non biodegradabili; inoltre hanno viscosità minore di quella dell'acqua e densità maggiore, per cui una volta immessi nell'ambiente tendono a scendere nel sottosuolo e ad accumularsi negli acquiferi sotterranei. Un'altra caratteristica importante è l'alta volatilità, per cui gran parte di essi evapora durante e dopo l'utilizzo (ad esempio nelle vernici).

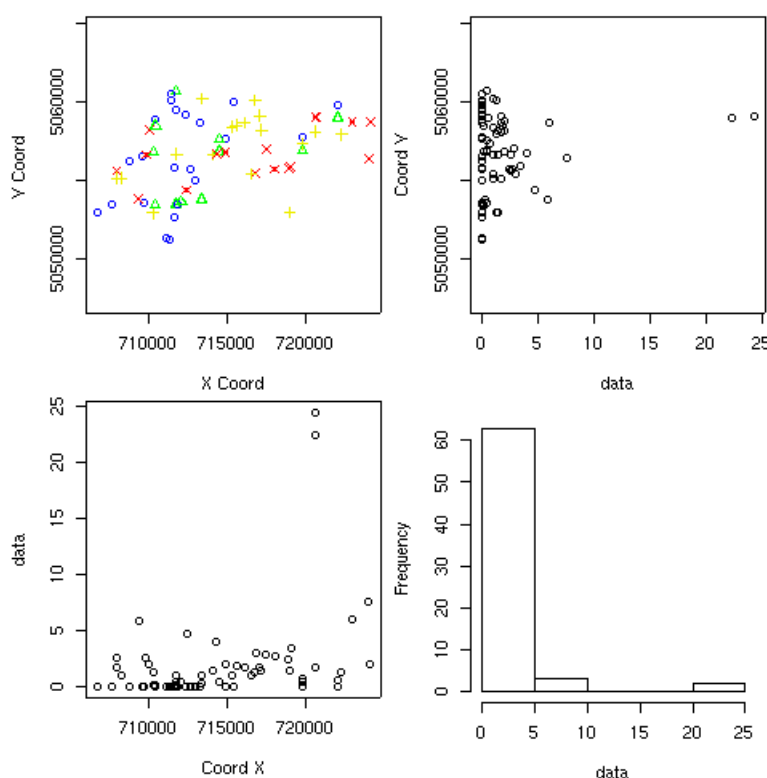


Figura 4.17: Disposizione dei pozzi e istogramma dei dati

Come si nota in figura 4.17 e nella tabella delle statistiche di sintesi, i Composti organo-alogenati si trovano in basse concentrazioni nell'area e hanno una distribuzione molto asimmetrica. Anche eliminando i due valori estremi, che si notano nell'istogramma, la distribuzione continua ad essere particolarmente asimmetrica. Non sembrano esserci direzioni preferenziali di sviluppo del fenomeno, caratteristica ragionevole per questo tipo di sostanze che, al contrario dei nitrati, derivano da fonti di inquinamento puntuale.

COMPOSTI ORGANO-ALOGENATI	
MINIMO	0.050 $\mu\text{g/l}$
PRIMO QUARTILE	0.100 $\mu\text{g/l}$
MEDIANA	0.650 $\mu\text{g/l}$
MEDIA	1.913 $\mu\text{g/l}$
TERZO QUARTILE	1.925 $\mu\text{g/l}$
MASSIMO	24.400 $\mu\text{g/l}$
DEVIAZIONE STANDARD	4.08998 $\mu\text{g/l}$
CURTOSI	23.49
ASIMMETRIA	4.42
Numero osservazioni 68	

Tabella 4.22: Statistiche di sintesi: misure di posizione, di variabilità e di forma

Si è deciso di ridurre l'influenza dei valori estremi utilizzando la trasformazione logaritmica. La distribuzione del logaritmo dei dati è evidentemente una mistura di Normali.

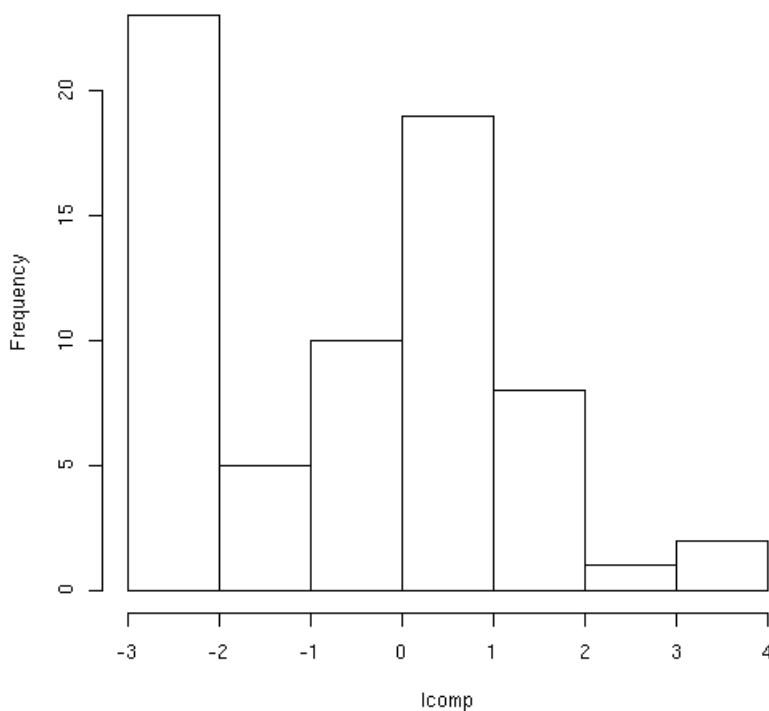


Figura 4.18: Istogramma dati trasformati

#### 4.4.2 Stima del variogramma

Nella seconda fase di analisi si procede con la stima del semivariogramma isotropo e si può notare dalla nuvola del variogramma che il fenomeno non si presenta regolare e continuo, poiché ci sono molti valori simili del variogramma stimato.

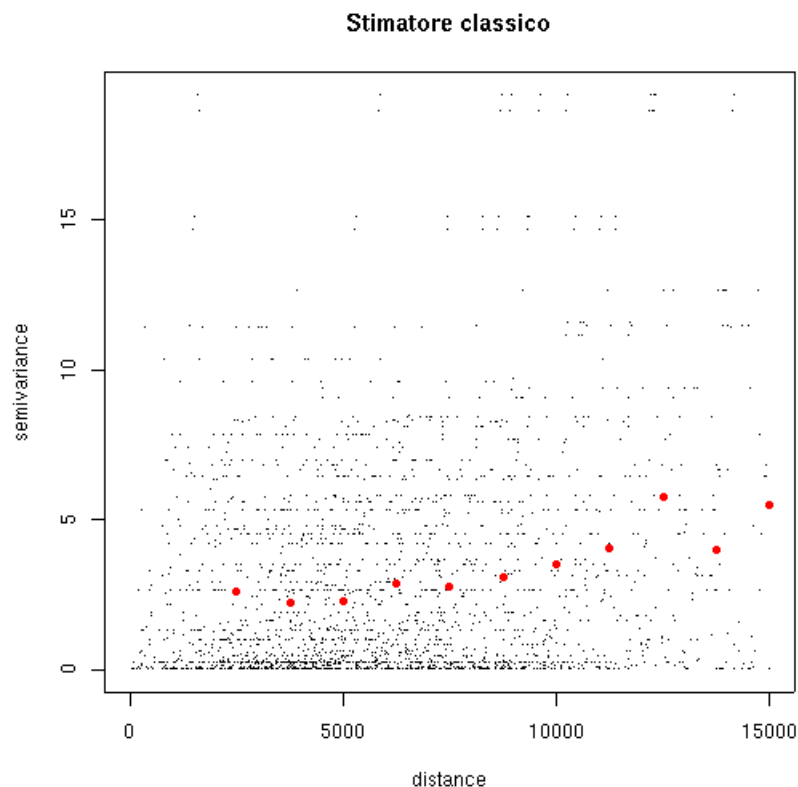


Figura 4.19: Nuvola del variogramma e stima del variogramma isotropo

Sebbene non siano state riscontrate direzioni preferenziali, si procede comunque con l'analisi dei semivariogrammi direzionali ( $\alpha = 0^\circ \pm 22.5^\circ$ ,  $\alpha = 45^\circ \pm 22.5^\circ$ ,  $\alpha = 90^\circ \pm 22.5^\circ$ ,  $\alpha = 135^\circ \pm 22.5^\circ$ ). Dall'analisi dei variogrammi direzionali non risultano direzioni preferenziali di sviluppo del fenomeno, come si era desunto dalla mappa di localizzazione. Per distanze inferiori ai 10 km il fenomeno può essere considerato isotropo, mentre per distanze superiori ai 10 km si ha maggiore variabilità.

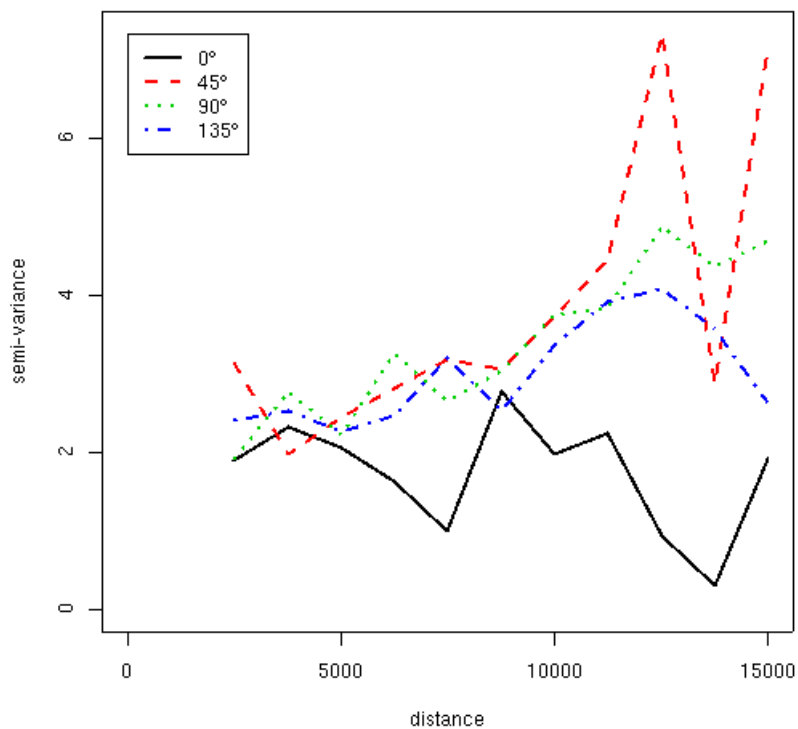


Figura 4.20: Variogrammi direzionali

#### 4.4.3 Costruzione del modello

Il modello teorico che si adatta in modo migliore al variogramma empirico è risultato essere quello esponenziale stimato con i minimi quadrati generalizzati (MQG).

	MODELLO ESPONENZIALE		
	M V	M Q G	M Q
<i>sill</i>	2.861	4.0259	6.853
<i>range</i>	476.6	3835.4682	11558.479
<i>nugget</i>	0	0	0



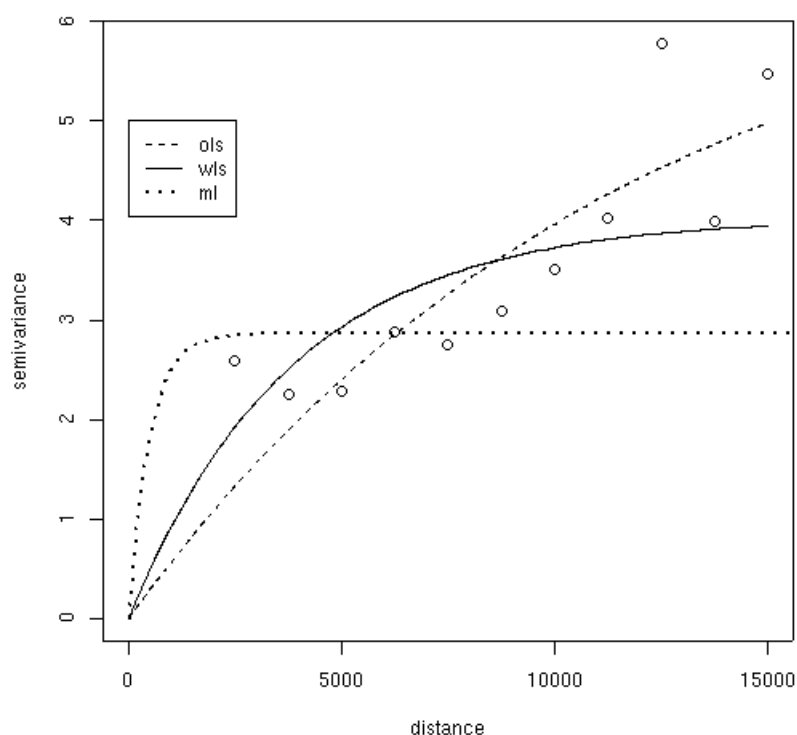


Figura 4.21: Stima dei parametri

#### 4.4.4 Previsione e rappresentazione grafica della concentrazione di Composti organo-alogenati

Per prevedere i valori dei Composti nella zona di interesse viene applicato il metodo del *kriging* ordinario, dal momento che non è stata riscontrata la presenza di un trend. Viene costruita una griglia regolare e sui nodi di questa vengono previsti i valori di concentrazione della sostanza (figura 4.22). Risulta evidente dalla mappa che l'inquinamento da composti ha origine in sversamenti puntuali, si notano infatti valori di concentrazione più elevata circondati da valori più bassi.

Di seguito si presenta la mappa della deviazione standard per esprimere l'affidabilità della previsione: è possibile notare elevata variabilità laddove ci sono pochi pozzi campionati.

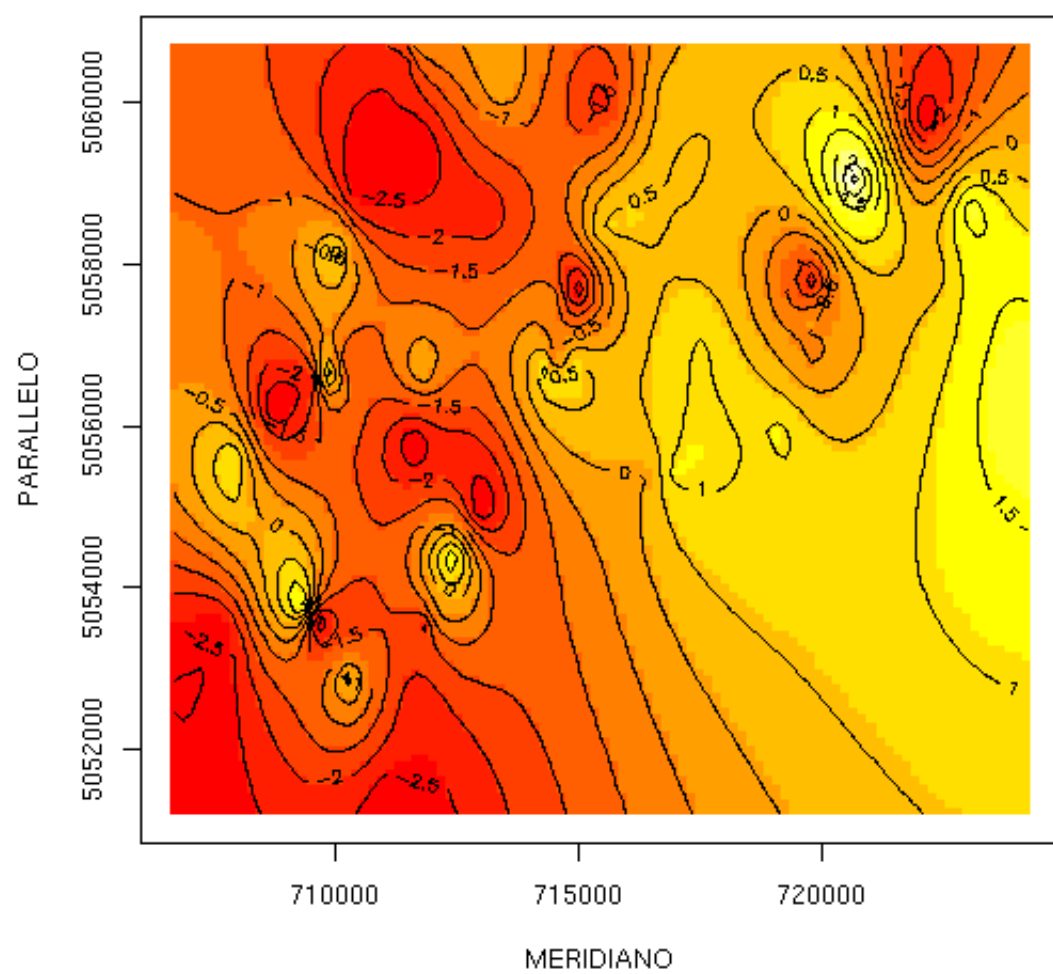


Figura 4.22: *Kriging* ordinario per i dati trasformati

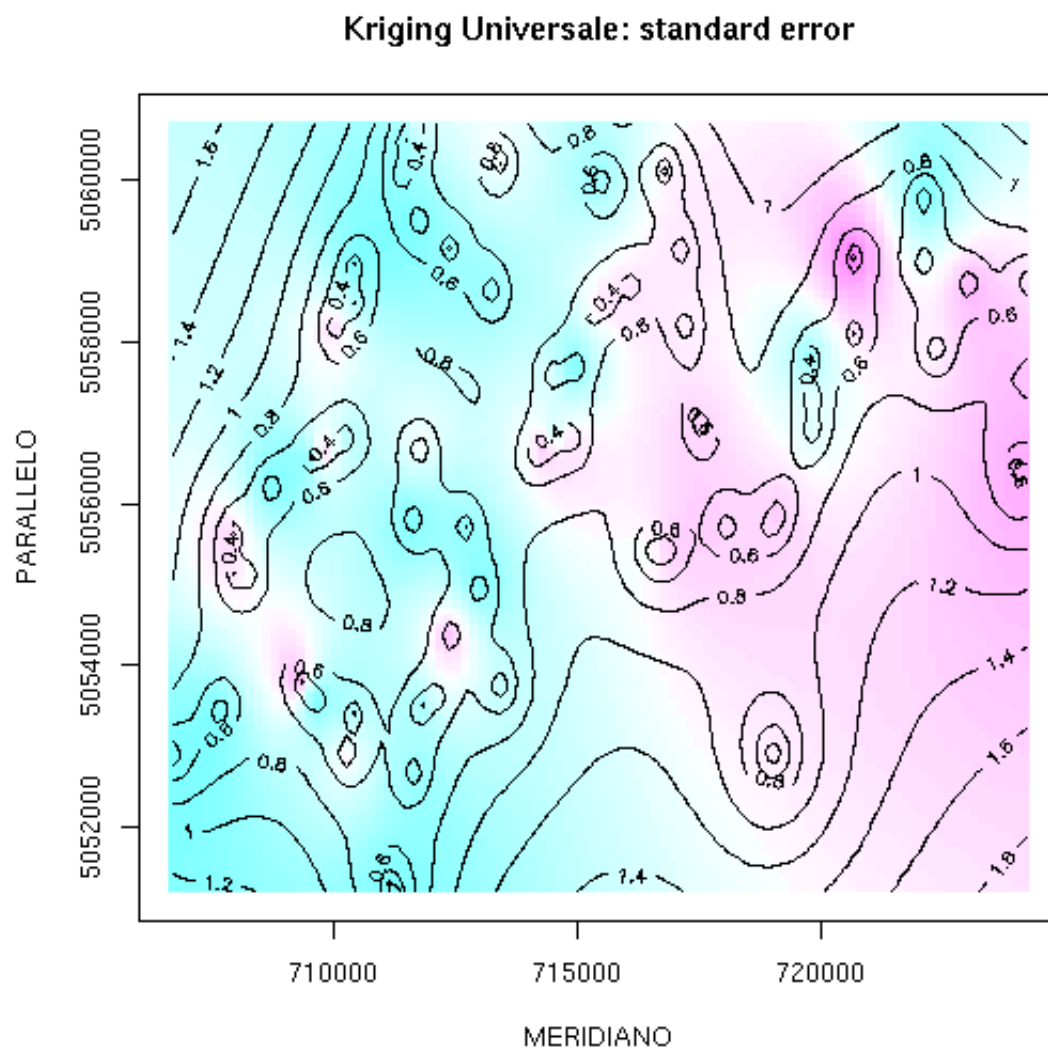


Figura 4.23: *Standard error* dei valori previsti di composti organo-alogenati

#### 4.4.5 Convalida del modello

La convalida del modello conduce ad un indice di *cross-validation*  $CV = 0.9872$ , dunque anche in questo caso si ottiene una buona previsione avendo applicato il *kriging* ordinario e il modello teorico di variogramma esponenziale.



# Conclusioni

La modellazione e previsione spaziale sono state utili per indagare eventuali situazioni di degrado nelle acque sotterranee.

La rappresentazione delle mappe di concentrazione ha permesso di notare come i cloruri presenti non siano inquinanti, viste le basse concentrazioni, ma abbiano origine naturale.

Le concentrazioni di nitrati presentano valori bassi, intorno ai 10 mg/l, con alcuni picchi di 20 mg/l, nella zona di San Pietro in Gù e valori elevati, fino a 45 mg/l, nei comuni di San Martino di Lupari, Tombolo e Cittadella. Lo ione nitrato è ritenuto un parametro significativo per quanto riguarda l'inquinamento diffuso. La presenza di molti insediamenti produttivi contribuisce ad aumentare il rischio di inquinamento della falda, che può essere determinato da sversamenti sul terreno, da scarichi delle fognature delle attività zootecniche. Questa è probabilmente la causa di concentrazioni così elevate di nitrati in pozzi a profondità minore di 30 metri.

Risulta evidente dalla mappa che l'inquinamento da composti ha origine in sversamenti puntuali, si notano infatti valori di concentrazione più elevata circondati da valori più bassi. La zona in cui si trovano Tombolo, Cittadella e San Martino di Lupari risulta essere quella maggiormente inquinata.



# Appendice A

## Rete di monitoraggio

Con deliberazione della Giunta Regionale n. 5571 del 17 ottobre 1986 è stato approvato il "Piano per il rilevamento delle caratteristiche qualitative e quantitative dei corpi idrici della Regione del Veneto (PRQA)", strumento che ha incorporato i vari programmi in precedenza predisposti per il controllo delle falde acquifere sotterranee.

La Regione, con DRG n. 3003/98, ha affidato all'Agenzia Regionale per la Prevenzione e Protezione Ambientale del Veneto (ARPAV) il compito di eseguire e coordinare l'attività di monitoraggio delle acque sotterranee del Veneto.

Nell'ambito del "Censimento dei corpi idrici e Piano Regionale di Risana-mento delle Acque (PRRA)" è stata predisposta una rete regionale di controllo delle acque sotterranee del Veneto.

Tale rete di monitoraggio è stata predisposta dal Dipartimento dell'Ambien-te della Regione Veneto nel 1983.

Per il monitoraggio qualitativo delle acque sotterranee, sono state eseguite, proprio a partire dal 1983, una serie di indagini sperimentali, non a cadenza periodica, sullo stato dell'inquinamento della falda freatica dell'Alta Pianura Veneta compresa tra Bassano e Cittadella. Sono stati utilizzati pozzi campio-nabili (idonei al prelievo di campioni) appartenenti alla rete di monitoraggio, ed anche opere di captazione di impianti pubblici che forniscono acque per uso potabile.

Per maggiori informazioni si veda [7].





## **Appendice B**

### **I dati analizzati**

## B.1 I 68 pozzi con profondità minore di 30 metri

Meridiano	Parallelo	Comune	Metri	A	B	C
711080	5051320	GAZZO	20.0	12.0	5.2	0.05
711740	5053440	GRANTORTO	6.5	6.2	14.7	0.20
711760	5056680	CARMIGNANO	18.0	7.2	20.2	1.00
719000	5052920	S. GIORGIO IN B.	18.0	11.4	29.9	1.50
711740	5059500	CITTADELLA	20.0	3.1	14.1	0.05
712380	5059160	CITTADELLA	25.0	5.8	16.4	0.10
712400	5054380	GRANTORTO	13.0	7.2	17.5	4.70
713400	5053760	GRANTORTO	15.0	7.8	7.7	0.30
714520	5056840	FONTANIVA	18.0	6.8	25.4	0.40
720660	5059020	TOMBOLO	13.0	14.3	30.4	22.40
720680	5058100	TOMBOLO	20.0	10.2	26.4	1.80
720680	5059060	TOMBOLO	20.0	13.2	32.0	24.40
711800	5053500	GRANTORTO	18.0	7.7	21.3	0.10
723020	5058720	S. MARTINO DI L.	12.0	11.6	34.0	6.00
724200	5058740	S. MARTINO DI L.	16.0	11.3	46.6	2.00
719790	5057800	TOMBOLO	14.0	13.9	33.6	0.10
722340	5057920	S. MARTINO DI L.	15.0	15.9	46.2	1.30
711140	5051260	GAZZO	14.0	14.4	3.2	0.05
717200	5058200	CITTADELLA	15.0	7.3	28.9	1.40
706700	5052900	S. PIETRO IN GU'	18.0	17.1	2.1	0.05
709620	5053580	S. PIETRO IN GU'	15.0	14.5	12.2	0.05
711300	5051220	GAZZO	8.0	17.2	2.1	0.05
711600	5055800	CARMIGNANO	13.0	7.9	14.6	0.05
712080	5053600	GRANTORTO	12.0	6.9	16.7	0.50
722100	5059000	S. MARTINO DI L.	15.0	12.9	32.4	0.60
711380	5060080	CITTADELLA	20.0	3.7	11.7	0.10
718060	5055720	CITTADELLA	17.0	7.5	27.4	2.80
713400	5060240	CITTADELLA	5.0	5.7	20.4	1.00
724100	5056400	S. MARTINO DI L.	21.0	10.6	36.1	7.60
707920	5055620	S. PIETRO IN GU'	7.0	6.6	17.3	2.60
712680	5055720	GRANTORTO	15.0	6.6	15.8	0.10
712980	5054960	GRANTORTO	10.0	8.1	14.0	0.05
717520	5057000	CITTADELLA	18.0	7.4	26.2	2.90
708240	5055120	S. PIETRO IN GU'	15.0	9.1	19.5	1.10
715320	5058400	CITTADELLA	25.0	5.7	21.0	1.10
716840	5055420	FONTANIVA	18.0	8.3	26.2	3.00

Meridiano	Parallelo	Comune	Metri	A	B	C
718960	5055740	CITTADELLA	20.0	13.8	29.0	2.50
714120	5056620	FONTANIVA	18.0	6.8	23.1	1.40
713220	5058640	FONTANIVA	18.0	2.9	2.2	0.10
716540	5055420	FONTANIVA	16.0	8.7	25.1	1.00
716780	5060120	CITTADELLA	28.0	5.9	24.5	1.30
709540	5056560	CARMIGNANO	14.0	6.5	16.6	0.10
711620	5052680	GRANTORTO	8.0	9.5	25.3	0.10
710420	5058940	CARMIGNANO	20.0	4.0	10.7	0.05
711700	5060690	CITTADELLA	25.0	3.9	8.9	0.50
714900	5056800	FONTANIVA	17.0	7.8	27.3	2.10
717100	5059140	CITTADELLA	20.0	7.3	25.8	1.80
716100	5058700	CITTADELLA	24.0	5.6	20.9	1.80
707660	5053440	S. PIETRO IN GU'	18.0	1.3	2.2	0.05
710260	5052900	S. PIETRO IN GU'	15.0	11.0	13.8	1.30
714960	5057700	FONTANIVA	15.0	5.6	18.1	0.05
719100	5055880	TOMBOLO	20.0	9.4	26.8	3.50
707960	5055120	S. PIETRO IN GU'	12.0	8.8	18.2	1.80
711440	5060500	CITTADELLA	18.0	3.8	7.6	0.10
715460	5060000	CITTADELLA	22.0	5.5	19.8	0.10
719820	5057300	TOMBOLO	17.0	5.2	8.2	0.70
709340	5053780	S. PIETRO IN GU'	12.0	14.0	23.5	5.90
714340	5056710	FONTANIVA	14.0	6.2	21.0	4.00
719820	5056900	TOMBOLO	22.0	12.4	34.8	0.50
722100	5059780	S. MARTINO DI L.	20.0	3.5	2.2	0.05
710020	5058200	CARMIGNANO	15.0	6.4	20.5	2.10
710260	5056820	CARMIGNANO	18.0	8.1	16.5	0.20
710440	5058480	CARMIGNANO	15.0	4.9	13.0	0.20
709820	5056640	CARMIGNANO	20.0	6.2	17.0	2.60
710400	5053400	S. PIETRO IN GU'	9.0	10.0	13.1	0.20
708720	5056200	S. PIETRO IN GU'	17.0	7.4	16.0	0.05
714500	5057600	FONTANIVA	22.0	6.4	19.5	0.50
715600	5058500	CITTADELLA	20.0	6.4	24.7	1.90

## B.2 I 49 pozzi con profondità compresa tra 30 e 60 metri

Meridiano	Parallelo	Comune	Metri	A	B	C
714240	5060620	CITTADELLA	32	4.0	8.8	0.30
716780	5060160	CITTADELLA	40	4.5	18.9	0.50
722240	5062120	S. MARTINO DI L.	30	8.9	36.6	3.10
719120	5052280	S. GIORGIO IN B.	60	1.7	2.1	0.05
718400	5052420	S. GIORGIO IN B.	56	1.6	1.2	0.05
719940	5051540	S. GIORGIO IN B.	58	1.9	3.0	0.05
719940	5052720	S. GIORGIO IN B.	60	2.0	3.0	0.05
725140	5054520	S. MARTINO DI L.	60	1.9	3.4	0.05
717840	5058320	CITTADELLA	60	2.7	4.9	0.05
718000	5053520	S. GIORGIO IN B.	50	2.0	2.7	0.10
721020	5055100	S. GIORGIO IN B.	50	5.1	7.4	0.80
719060	5052980	S. GIORGIO IN B.	60	2.3	3.0	0.05
713920	5054280	GRANTORTO	45	1.3	0.5	0.05
715220	5057920	FONTANIVA	51	2.5	3.8	0.10
718600	5053740	S. GIORGIO IN B.	50	2.3	3.0	0.05
720900	5054620	S. GIORGIO IN B.	58	3.4	5.3	0.60
720120	5054180	S. GIORGIO IN B.	55	3.1	6.0	0.40
720840	5061380	GALLIERA	50	9.8	35.6	1.90
718220	5051840	S. GIORGIO IN B.	54	1.0	0.7	0.05
720200	5049660	GALLIERA	60	2.1	1.1	0.05
716960	5056460	CITTADELLA	60	2.1	4.3	0.05
721180	5054400	S. GIORGIO IN B.	60	4.2	5.8	0.70
714760	5056460	FONTANIVA	60	1.9	1.1	0.05
715180	5057120	FONTANIVA	52	5.9	20.0	0.05
721960	5058560	S. MARTINO DI L.	40	14.8	34.4	3.10
722920	5058860	S. MARTINO DI L.	30	14.2	47.4	3.70
715960	5058360	CITTADELLA	52	2.9	4.5	0.05
723140	5060080	S. MARTINO DI L.	52	3.6	8.8	0.90

- A: Cloruri mg/l
- B: Nitrati mg/l
- C: Composti Organo-alogenati  $\mu\text{g/l}$
- Metri: profondità in metri

Meridiano	Parallelo	Comune	Metri	A	B	C
718040	5051820	S. GIORGIO IN B.	60	1.4	0.5	0.05
719200	5055080	CITTADELLA	55	2.7	4.2	0.05
718920	5052180	GALLIERA	54	1.8	2.1	0.05
716420	5060810	CITTADELLA	57	3.0	5.0	0.20
718540	5051100	S. GIORGIO IN B.	40	2.7	0.5	0.30
719800	5048500	C. S. MARTINO	53	1.1	1.3	0.05
720880	5059180	TOMBOLO	30	4.5	8.0	1.50
721140	5055380	S. GIORGIO IN B.	52	13.7	36.8	1.50
717620	5059120	CITTADELLA	55	3.2	4.7	0.10
719600	5050380	S. GIORGIO IN B.	60	2.1	2.6	0.05
719880	5056400	TOMBOLO	56	13.6	36.0	0.70
722500	5056040	S. MARTINO DI L.	42	2.5	3.5	0.05
725080	5054680	S. MARTINO DI L.	60	2.0	2.9	0.05
714420	5056680	FONTANIVA	57	2.1	1.4	0.05
715620	5054940	FONTANIVA	42	1.1	0.5	0.05
716620	5056900	CITTADELLA	60	2.2	3.8	0.05
719980	5052560	S. GIORGIO IN B.	60	1.9	2.3	0.05
712120	5056520	CARMIGNANO	50	1.8	2.4	1.60
714060	5053180	GRANTORTO	47	0.8	0.5	0.10



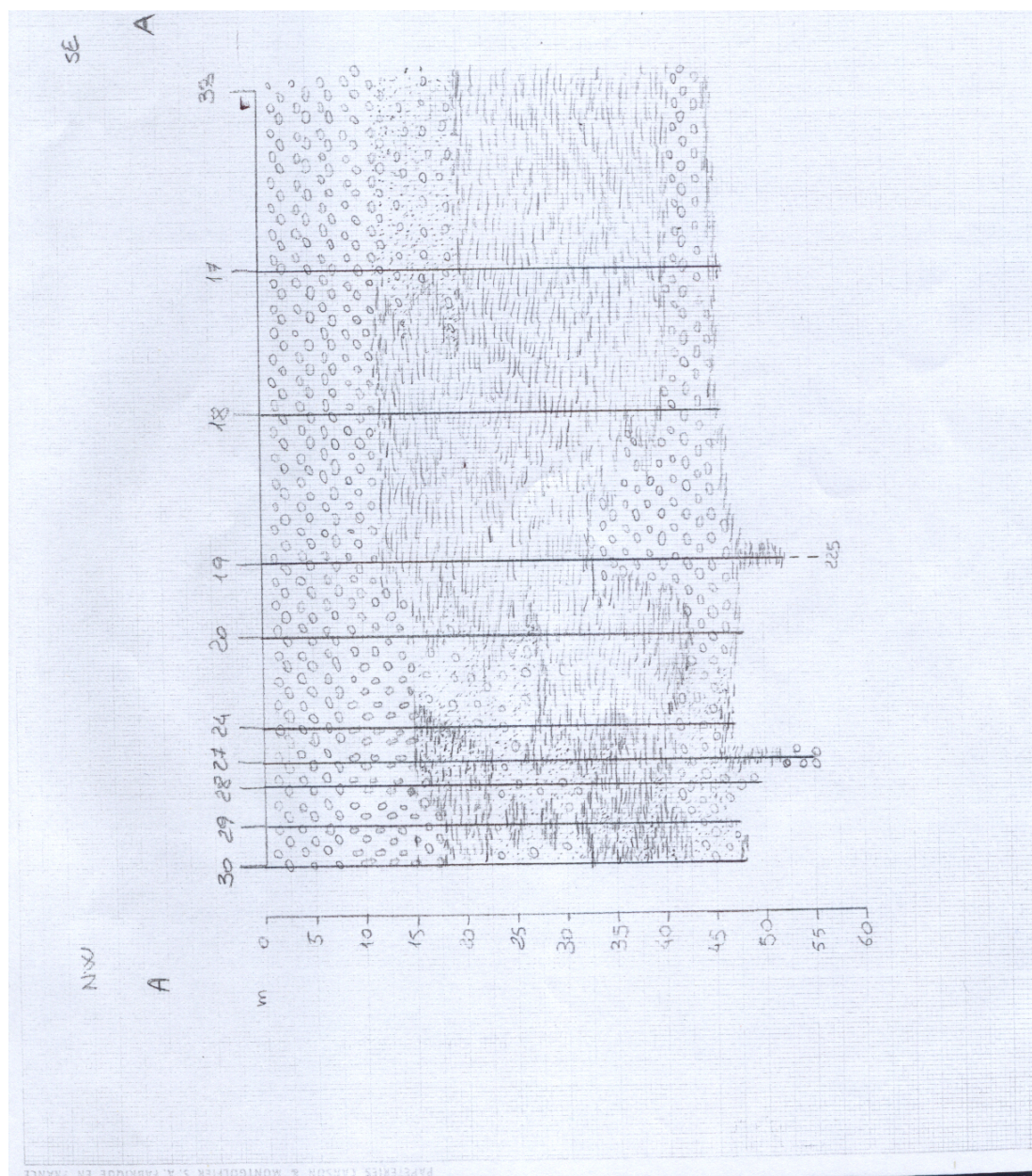
## **Appendice C**

### **Zona di interesse e stratigrafie**

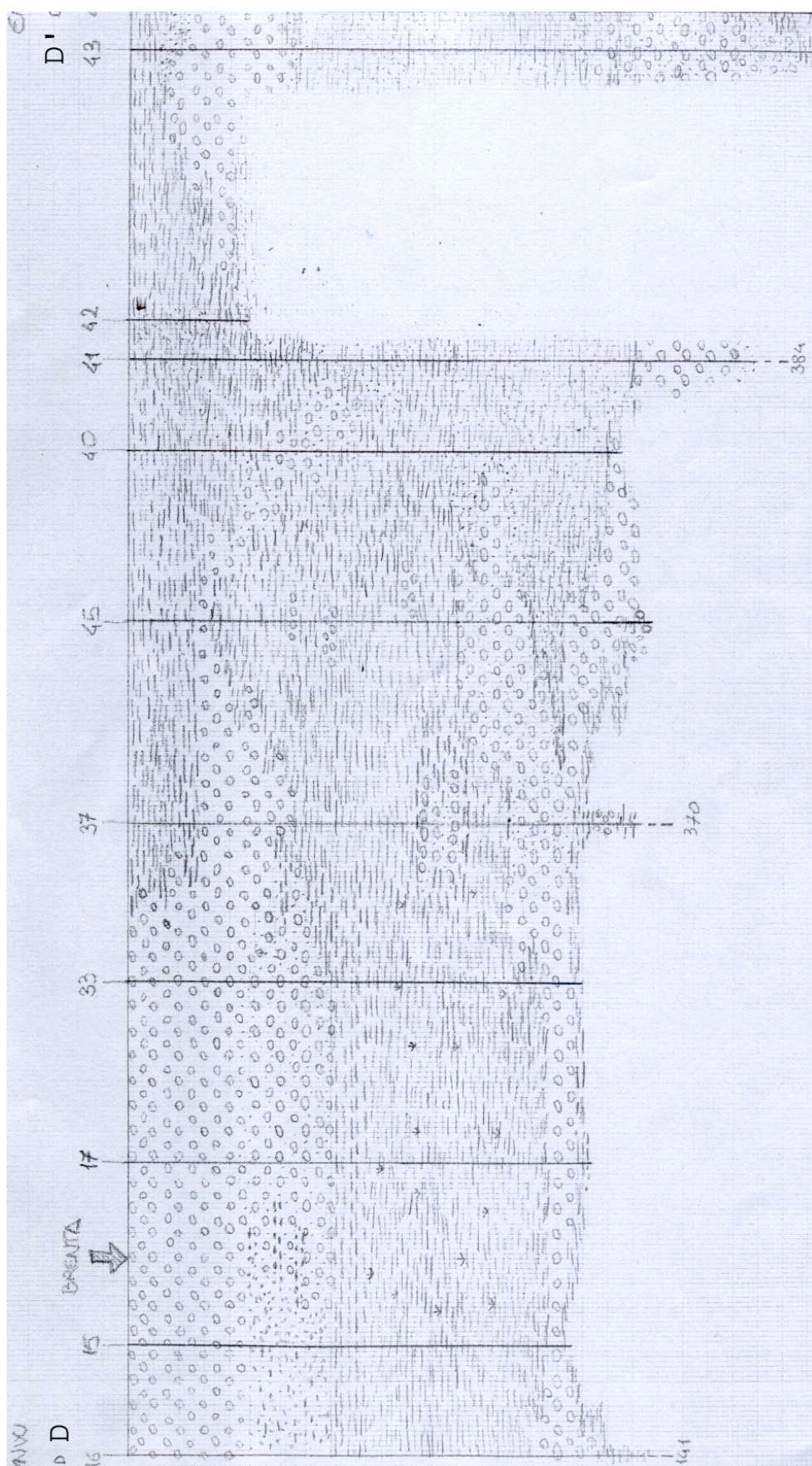




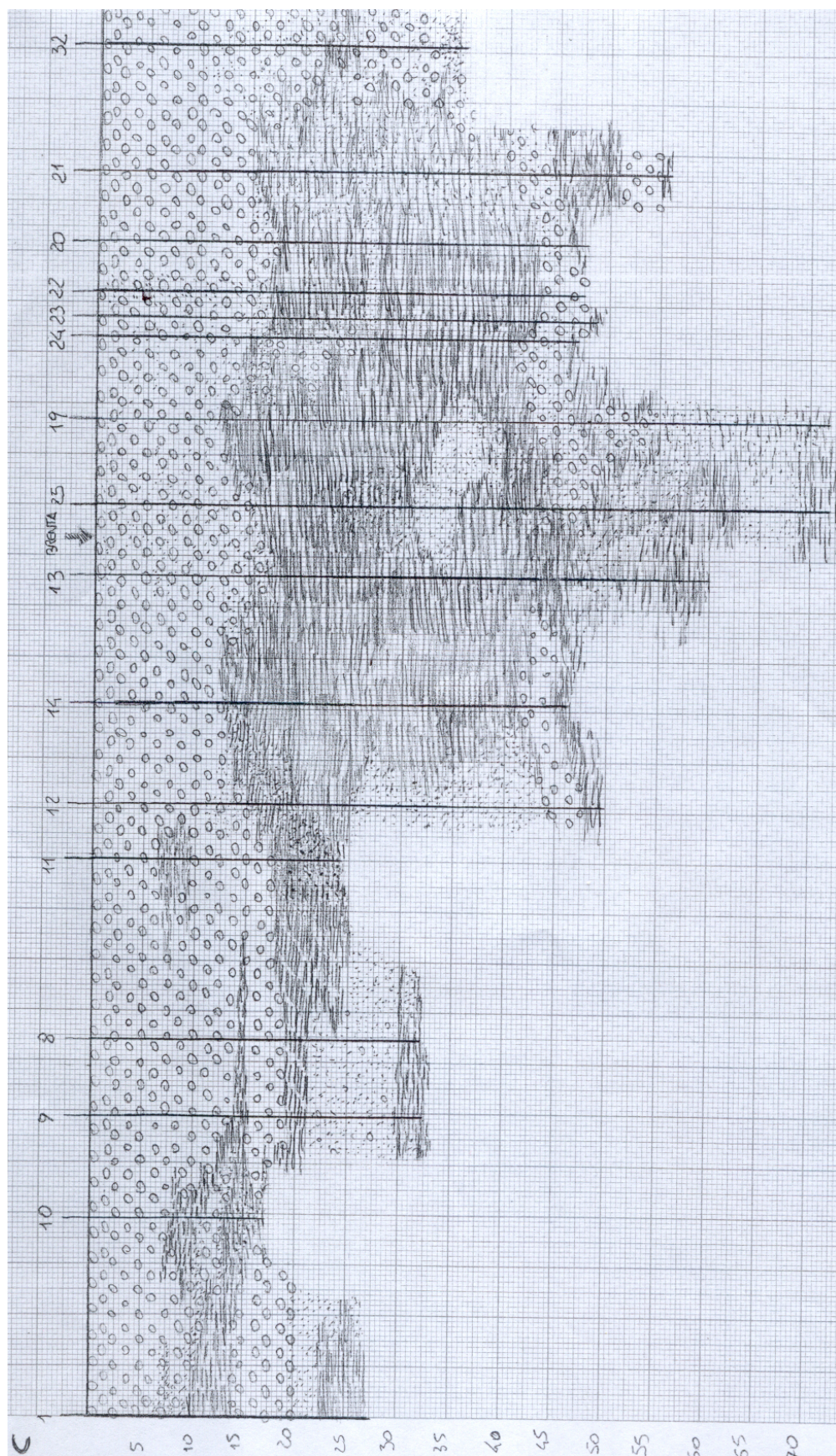














# Appendice D

## Il *software* R

Il software R è un linguaggio di programmazione per l'analisi statistica che permette di elaborare grandi quantità di dati, di eseguire calcoli complessi, di effettuare rappresentazioni grafiche e che può essere esteso dall'utente finale.

E' stato implementato da R. Ihaka e R. Gentleman del Dipartimento di Statistica dell'Università di Auckland in Nuova Zelanda e dal 1997 il suo sviluppo è affidato al gruppo internazionale *R Development Core Team*.

Spesso questo linguaggio viene denominato ambiente integrato piuttosto che software poiché è costruito a partire dagli oggetti (tabelle, grafici, vettori, insiemi di dati...) che vengono trattati con metodi specifici attraverso le funzioni di R.

Caratteristiche peculiari di R sono l'essere interamente gratuito e *open source*, infatti tutti possono accedere al codice interno e proporre modifiche ed è disponibile secondo la licenza GPL (*Gnu Public Licence*) all'indirizzo internet <http://cran.r-project.org/>.

Per le applicazioni sono stati utilizzati i *package* geoR e rgl.



# Bibliografia

- [1] Cressie N. (1993) *Statistics for Spatial Data. 2nd edition*, Wiley, New York
- [2] Fabbri L. (1997) *Statistica multivariata - Analisi esplorativa dei dati*, McGraw-Hill, Milano
- [3] Goovaerts P. (1997) *Geostatistics for Natural Resources Evaluation*, Oxford University Press, Oxford
- [4] Isaaks E. H. - Srivastava R. M. (1989) *Applied Geostatistics*, Oxford University Press, Oxford
- [5] Journel A. G. - Huijbregts C. J. (1981) *Mining geostatistics*, Academic Press, London
- [6] Kitanidis P. K. (1997) *Introduction to geostatistics*, Cambridge University Press
- [7] Mion F. (2002) *Acqua sotterranea nel Veneto*, Osservatorio Regionale Acque-Area tecnico scientifica, ARPA VENETO, [www.arpa.veneto.it](http://www.arpa.veneto.it)
- [8] Pebesma E. J. , de Kwaadsteniet J. W. (1997) *Mapping groundwater quality in the Netherlands*, Journal of Hydrology, pag. 364-386
- [9] Posa, D. , (1995) *Introduzione alla Geostatistica*. Adriatica Editrice Salentina, Lecce
- [10] Wackernagel H. (2003) *Multivariate Geostatistics*, Springer-Verlag, Berlin
- [11] Zanghieri P. (2000) *L'acqua sotterranea: una risorsa nascosta - Pozzi, acquiferi e falde nella provincia di Venezia* , Assessorato alle Politiche Ambientali della Provincia di Venezia, [www.provincia.venezia.it](http://www.provincia.venezia.it)